

STIC-Biotech/ChemLib

79281

From: Ibrahim, Medina A.  
Sent: Friday, November 01, 2002 8:29 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/905, 558

Please search the following:

1. SEQ ID NO:3 and 16. Please search both commercial and issued patents databases. Thanks

Medina A. Ibrahim  
Patent Examiner  
GAU:1638  
CM1-9A12  
mailbox9E12  
(703)306-5822

RECEIVED  
NOV-4 2002  
STIC

Mary Jane Ruhl  
Tech. Info. Specialist, STIC  
TC-1600  
CM-1, Room 6A-06  
Phone: 605-1155

Mary Jane Ruhl  
Tech. Info. Specialist, STIC  
TC-1600  
CM-1, Room 6A-06  
Phone: 605-1155

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

10

1

10

10

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

***The Pending database search results should not be left in the case because they contain data that is confidential.***

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

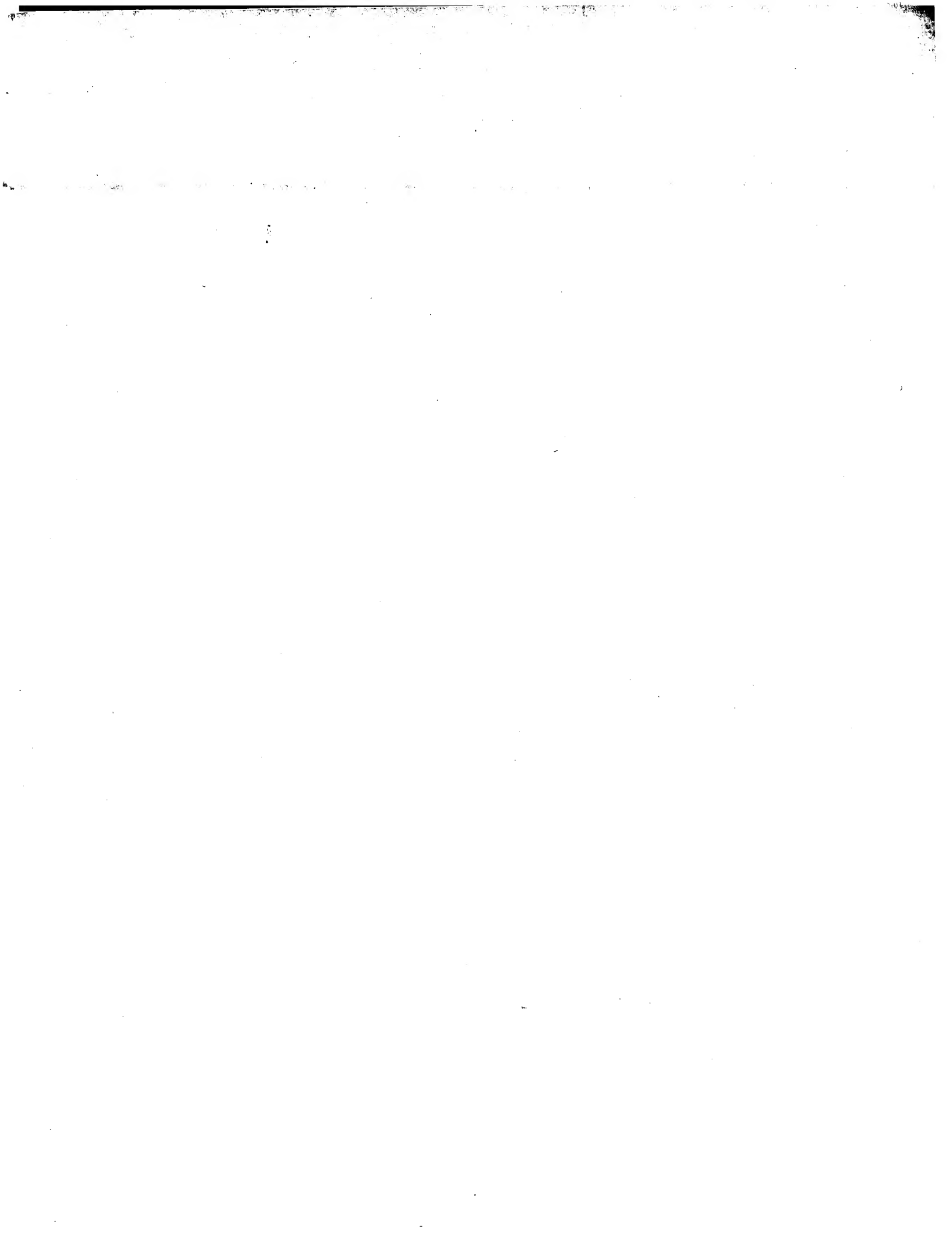
Two new databases have been created to hold the pre-published sequences:

**Published\_Applications\_NA** contains nucleic acid sequences; the search results will have the extension **.rnpb**.

**Published\_Applications\_AA** contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

**Sequences in the PGPub database are public information; it is permissible to leave these results in the case.**





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:47:38 : Search time 1687.14 Seconds  
(without alignments)  
16248.621 Million cell updates/sec

Title: US-09-905-558c-3  
Perfect score: 1310  
Sequence: 1 cccatcgctgtttgtctac.....caagggaagtgatcgatg 1310

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_Other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----	-----	-----	-----	-----	-----

1	232.8	17.8	2126	6	AX299951	Sequence
2	124.6	9.5	1888	6	AX099709	Sequence
3	112.2	8.6	9285	8	ZMU09989	Zea mays D3
4	60.2	4.6	174802	2	AC108209	Homo sapi
5	58.2	4.4	125020	9	AF429315	Homo sapi
6	56.6	4.3	173848	9	AC090043	Homo sapi
7	56.2	4.3	169660	9	AL161449	Human DNA
8	55.6	4.2	204652	2	PFMAL13P6	Plasmid
9	55.2	4.2	650	3	SGU36751	Schizaphis
10	54.8	4.2	204652	2	PFMAL13P6	Plasmid
11	54	4.1	89904	8	ATW25K17	Arabidops
12	54	4.1	196286	8	ATCHRIV64	Arabidops
13	53.8	4.1	115522	8	AC006248	Arabidops
14	53.6	4.1	214033	2	AC105576	Rattus no
15	53.4	4.1	614	3	AF254090	Rhopalosi
16	53.2	4.1	167553	9	AC090042	Homo sapi
17	53	4.0	127539	9	CNS01DSN	Human chr
18	53	4.0	146868	2	AC016802	Homo sapi
19	53	4.0	151117	2	AL357128	Homo sapi
20	53	4.0	161277	2	AC091953	Homo sapi
21	53	4.0	176033	2	AL161794	Homo sapi
22	53	4.0	178342	9	AC007435	Homo sapi
23	53	4.0	187150	2	AC025924	Homo sapi
24	53	4.0	199385	2	AC019049	Homo sapi
25	52.6	4.0	72090	2	AC095644	Rattus no
26	52.6	4.0	80692	2	AC100569	Mus muscu
27	52.6	4.0	110000	2	PFMAL4P1_0	Plasmid
28	52.6	4.0	191772	2	AC078975	Plasmid
29	52.6	4.0	293431	2	PFMAL13P4	Plasmid
30	52.4	4.0	171574	2	AC012300	Homo sapi
31	52.4	4.0	196203	9	AL355355	Human DNA
32	52.2	4.0	90487	9	AL592166	Human DNA
33	52.2	4.0	126801	9	AC004849	Homo sapi
34	52.2	4.0	149450	2	AC067880	Homo sapi
35	52.2	4.0	153300	9	AC073387	Homo sapi
36	52.2	4.0	170141	2	AL611928	Homo sapi
37	52.2	4.0	213692	2	AL590427	Homo sapi
38	52	4.0	159593	9	AC004832	Homo sapi
39	51.8	4.0	678	3	AF143595	Myzus var
40	51.8	4.0	14135	3	AE001388	Plasmid
41	51.6	3.9	79646	8	ATT10D17	Arabidops
42	51.6	3.9	162996	9	AC006441	Arabidops
43	51.4	3.9	179307	2	AC104027	Homo sapi
44	51.2	3.9	859	8	AF292708	Homo sapi
45	51.2	3.9	7218	6	I66494	Sequence 14

ALIGNMENTS

RESULT 1	AX299951	Sequence	2126 bp	DNA	linear	PAT 26-NOV-2001
LOCUS	AX299951	Sequence	92 from Patent WO0183790.			
DEFINITION	AX299951	Sequence	92 from Patent WO0183790.			
ACCESSION	AX299951	Sequence	92 from Patent WO0183790.			
VERSION	AX299951.1	GI:17129442				
KEYWORDS						
SOURCE	Zea mays.					
ORGANISM	Zea mays					
REFERENCE	1 (sites)					
AUTHORS	Conner,T.W., Dubois,P., Malven,M. and Masucci,J.D.					
TITLE	Plant regulatory sequences for selective control of gene expression					
JOURNAL	Patent: WO 0183790-A 92 08-NOV-2001;					
FEATURES	Monsanto Technology LLC (US)					
Source	Location/Qualifiers					
	1..2126					
	/organism="Zea mays"					
	/db_xref="taxon:4577"					
BASE COUNT	607 a	419 c	433 g	667 t		
ORIGIN						



exon	3927..4031	/gene="Mhal"
	/number=11	
intron	4032..4112	/gene="Mhal"
	/number=11	
exon	4113..4257	/gene="Mhal"
	/number=12	
intron	4258..4378	/gene="Mhal"
	/number=12	
exon	4379..4543	/gene="Mhal"
	/number=13	
intron	4544..4636	/gene="Mhal"
	/number=13	
exon	4637..4875	/gene="Mhal"
	/number=14	
intron	4876..4963	/gene="Mhal"
	/number=14	
exon	4964..4996	/gene="Mhal"
	/number=15	
intron	4997..5090	/gene="Mhal"
	/number=15	
exon	5091..5172	/gene="Mhal"
	/number=16	
intron	5173..5256	/gene="Mhal"
	/number=16	
exon	5257..5417	/gene="Mhal"
	/number=17	
intron	5418..5512	/gene="Mhal"
	/number=17	
exon	5513..5689	/gene="Mhal"
	/number=18	
intron	5690..5767	/gene="Mhal"
	/number=18	
exon	5768..5950	/gene="Mhal"
	/number=19	
intron	5951..6035	/gene="Mhal"
	/number=19	
exon	6036..One-of(6408,6409)	
	/number=20	
	/evidence=experimental	
polyA_site	6408	/evidence=experimental
	6409	/evidence=experimental
polyA_site	6409	/evidence=experimental
BASE COUNT	2329 a 2181 c 2086 g 2689 t	
ORIGIN		
Query Match	8.6%;	Score 112.2; DB 8; Length 9285;
Best Local Similarity	74.1%;	Pred. No. 8.8e-16;
Matches 177; Conservative	0; Mismatches 43; Indels 19; Gaps	
Qy 78	CCGTTTCGAGTCACATGGCGTACACTTTCTGCATCGACTTTGTACGGCTACATCGAA	137
Db 220	CCGTTCCAGGAGTCGACTGTGTACACCTGTCTGCATCGAATTGTGTACTACTCGAA	161
Oy 138	CATATACACGAGATGTCTCGTGTGTAATAGAGTCACTAATGCCCTTAAGCATCGGTTACTCC	197



TITLE A repeat expansion in the gene encoding juncctophilin-3 is associated with Huntington disease-like 2  
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)  
MEDLINE 21583737  
PUBMED 11694876  
REFERENCE 2 (bases 1 to 125020)  
HOLMES,S.E., INGERSOLL,Ashworth,R.G., ROSS,C.A. and MARGOLIS,R.L.  
Direct Submission  
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
FEATURES  
source Location/Qualifiers  
1..125020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3; between D16S520 and WI-12410"  
/note="Isolated from a patient with Huntington's Disease-Like 2 (HDL2)"  
repeat\_region complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit=ctg  
mRNA complement(<36507..>36887)  
/gene="JPH3"  
/product="Juncctophilin 3"  
gene complement(<36507..>36887)  
/gene="JPH3"  
/note="Jp3"  
CDS complement(<36507..36887)  
/gene="JPH3"  
/note="component of the functional complex between plasma membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="Juncctophilin 3"  
/protein\_id="AA140941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGRNFDGSGYCGWBDGKAHGHVCTGPKGQGEYTGSMW HGFVLGYTPNSDTGCTGTAQCKRIGLESKGVYKGEWTHGFKRGYVRECA  
BASE\_COUNT 29056 a 32731 c 30696 g 28283 t 4254 others  
ORIGIN  
Query Match 4.4%; Score 58.2; DB 9; Length 125020;  
Best Local Similarity 10.6%; Pred. No. 0.0096;  
Matches 60; Conservative 248; Mismatches 256; Indels 0; Gaps 0;  
QY 745 GCACGAGGGCACTACCCCACTCTCACGAAACCGCGCTGATCGGCAATCAAC 804  
Db 16892 KSWMSBMSVSYSVKHMSHBSCHWBKMTWSCCMYKSWGSGWGMCCWGR 16951  
QY 805 GAGGTGGTGGCCCGTCCAGCTCCAGGCGGACCATCCCTCTGCAGCGGCTCAC 864  
Db 16952 SKGWKWSRGSRSKSMRYTGGSKMRSSMMCTSCYASMCWCCWCCMRSCC 17011  
QY 865 CAGCATCCGCTGTCGGGAACGACACACACCCCAACCATCTCAGCAACCCGCTCC 924  
Db 17012 YCCMRYCCACKYNSWTSYMSRYSYKRSYKRSYKRSYKRSYKRSYKRSY 17071  
QY 925 CGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 984  
Db 17072 GKGKGGYGRKTRKSRGKMGAKWYMYRSMRKMWYSSKGMWYCWGRRGRCYC 17131  
QY 985 GGACACCGGACACCTGCGGCCCTTTGTTATTCATCCGGAATCATCTGCCGCC 1044  
Db 17132 SCMTSRAMCCSYAKCKSNCSYCYTGMKGYACSYRGSMSKSYCMRGSTYSF 17191  
QY 1045 GGCGGACTGCGCTGCGCGCGCGGATATATACCCATCGTTATCGATCATGATCG 1104  
Db 17192 CCCTTTTCCTCCNANTGGGAAGCTTTNCKNTSYIRKRNCGAMCKYNNYNS 17251  
QY 1105 TCATCTACGGGTAGCTCATGTCGAGCGTAGCATGCAAGCACTATTTCGCGTCC 1164  
Db 17252 SCRAMGCTKYKSSMTSMASMSYKSWMSYKSWMSYKSWMSYKSWMSYKSW 17311

QY 1165 CAGGTCTCCGTCGCGTCCCTTCCAGTCTGTCTCACACTAGCTGCTGGGACGATGAA 1224  
Db 17312 KCCWCMKCYCMRSMRSGMSYMYASWSSRGCTCTRCYCMWSSKCYKSYMMRS 17371  
QY 1225 GTGGTGTCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 1284  
Db 17372 KRMGMKWSRCSGWSGMSASRSSCYKYSKMRCSMMSKCYKCAACGMMGGMRYMRCW 17431  
QY 1285 GCTGATCAAGGAAAGTATCGGA 1308  
Db 17432 SMKRRWKGKSAMYRMRMRWKGGA 17455

RESULT 6  
AC090043 173848 bp DNA linear PRI 11-FEB-2001  
LOCUS Homo sapiens chromosome 3 clone RP11-551L4 map 3p, complete  
DEFINITION AC090043  
AC090043  
VERSION AC090043.1 GI:12745081  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 173848)  
AUTHORS Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y., Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J., Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G., Chen,Z. and Huang M.

TITLE Chromosome 3p genomic sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 173848)  
AUTHORS Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,X., Li,W., Li,W., Li,X., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,X., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang H.

TITLE Direct Submission  
JOURNAL Submitted (11-FEB-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

COMMENT  
Center:Beijing Center  
Center code:Beijing  
Website:http://hgsc.igtp.ac.cn  
http://www.genomics.org.cn  
Contact:hgsc@igtp.ac.cn  
----- Project Information  
Center project name:1g project  
Center clone name: RP11-551L4  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator: ET 55% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 501 bases at least Q40  
Consensus quality: 829 bases at least Q30  
Insert size: 1157; sum-of-contigs  
Quality coverage: 1.35x in Q20 bases;sum-of-contigs

FEATURES  
source Location/Qualifiers  
1..173848  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"

on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
This sequence is the entire insert of clone RP11-382H24 The true  
left end of clone RP11-664D14 is at 135950 in this sequence. The  
true right end of clone RP11-187K14 is at 53851 in this sequence.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP11-382H24 is from  
the library RPCI-11.2 constructed by the group of Pieter de Jong.  
For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6.

FEATURES		Location/Qualifiers	
Source		1. 169660	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="9"	
		/map="p22.1-1-23"	
		/clone="RP11-382H24"	
		/clone_lib="RPCI-11.2"	
		2. 492	
misc_feature		/note="match: GSS: Em:AQ700717"	
misc_feature		complement(1479..2009)	
repeat_region		/note="match: GSS: Em:AQ569033"	
repeat_region		1589..1778	
repeat_region		/note="MIR repeat: matches 23..192 of consensus"	
repeat_region		1669..1724	
repeat_region		/note="2 copies 28 mer 92% conserved"	
repeat_region		2068..2401	
repeat_region		/note="L1MC4 repeat: matches 7453..7834 of consensus"	
repeat_region		2409..2828	
repeat_region		/note="15 copies 28 mer 62% conserved"	
repeat_region		2429..2848	
repeat_region		/note="7 copies 60 mer 65% conserved"	
repeat_region		2452..2801	
repeat_region		/note="175 copies 2 mer ta 61% conserved"	
repeat_region		2819..2911	
misc_feature		/note="L1MC4 repeat: matches 7864..7961 of consensus"	
misc_feature		complement(4102..4478)	
misc_feature		/note="match: GSS: Em:AQ180899"	
repeat_region		4497..5002	
repeat_region		/note="match: GSS: Em:AQ716496"	
repeat_region		4621..4865	
repeat_region		/note="MIR repeat: matches 1..261 of consensus"	
repeat_region		5200..5294	
repeat_region		/note="L1 repeat: matches 2578..2670 of consensus"	
repeat_region		6031..6654	
repeat_region		/note="L1PA12 repeat: matches 5583..6152 of consensus"	
repeat_region		7419..7966	
repeat_region		/note="match: GSS: Em:AQ378047"	
repeat_region		7431..7737	
repeat_region		/note="AluSc repeat: matches 1..307 of consensus"	
repeat_region		7763..8103	
repeat_region		/note="MLT1A1 repeat: matches 205..536 of consensus"	
repeat_region		8379..8422	
repeat_region		/note="22 copies 2 mer tg 86% conserved"	
repeat_region		9171..9447	
repeat_region		/note="L1M4 repeat: matches 2900..3176 of consensus"	
repeat_region		9450..9739	
repeat_region		/note="L1M4 repeat: matches 3869..4175 of consensus"	
repeat_region		9774..9835	
repeat_region		/note="L1M7 repeat: matches 6223..6284 of consensus"	
repeat_region		L1MA7 repeat: matches 6223..6284 of consensus	
repeat_region		10099..10533	
repeat_region		/note="MLT1C repeat: matches 9..466 of consensus"	

BASE COUNT		56663 a 31451 c 31882 g 53852 t	
ORIGIN		/map="3p"	
		/clone="RP11-551L4"	
Query Match		4.3%; Score 56.6; DB 9; Length 173848;	
Best Local Similarity		47.0%; Pred. No. 0.025;	
Matches 208; Conservative		0; Mismatches 234; Indels 1; Gaps 1;	
QY 100		TACATCTTTCTGCATCGACTTTGTAGGCTACATCGACATATACACGAGATGCTCTGTG 159	
Db 163709		TATAAATTTAT 163768	
QY 160		TGAATAGACTACTAATGCTTAAGCATCGGTACTCCGTAGGTACATTCGTCTTCT 219	
Db 163769		TACATATAAATAGATATATTTATATATATATATATATATATATATATATATATATAT 163828	
QY 220		TATTTCTGCATATTTTATTTGTTTACTGATTATACAGTAGTATACATACATGCAC 279	
Db 163829		TGTATTTAT 163888	
QY 280		ATACATATACATATATATACATATATTTTCTAAATTAATTAACCTAAATGACT 339	
Db 163889		ATGCATATTTATACATATATATTTATATATATATATATATATATATATATATATAT 163948	
QY 340		AAATTTCTAACCAACGACATTTGTAATGTTTCTCCCAACACTTTACCTAT-TCTACAT 398	
Db 163949		TTATACATATGTATGATATATTTATATATATATATATATATATATATATATATATAT 164008	
QY 399		TGTTCTATTTTCGAATTTCTACTCTATAAACAACATAGTCTCAATGGAACAGTCTTTG 458	
Db 164009		TATTTAT 164068	
QY 459		TAGACTATATACCGATGTGCGCTACACATACACATATAGTCGTTTGAAGATTGA 518	
Db 164069		TGTACAT 164128	
QY 519		ACCTATATATCGTACGGTTAAT 541	
Db 164129		ATGTACATATACATATATATAT 164151	
RESULT 7			
AL161449		169660 bp DNA linear PRI 04-DEC-2001	
LOCUS		Human DNA sequence from clone RP11-382H24 on chromosome 9p22.1-23	
DEFINITION		Contains a thiredoxin peroxidase pseudogene, a SSB (Sjogren syndrome antigen B (autoantigen La) pseudogene, the 3' end of the MPDZ gene for multiple PDZ domain protein and a CpG island, complete sequence.	
ACCESSION		AL161449	
VERSION		AL161449.7 GI:8894259	
KEYWORDS		HTG; CpG island; MPDZ.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		Sehra, H.	
JOURNAL		Direct Submission	
COMMENT		Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk On Jul 1, 2000 this sequence version replaced gi:8653807. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM:, EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information	







	a	c	g	t	others
BASE COUNT	68880	17396	17710	67856	32810
ORIGIN					

Query Match	4.28;	Score 54.8;	DB 2;	Length 204652;
Best Local Similarity	51.28;	Score 54.8;	DB 2;	Length 204652;

Best local similarity 51.3%; Pred. NO. 0.069;  
Matches 153; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

Qy 173 TAATGCTTAAAGCATCGGTACTCCGTAGGTACATTCGTCTCTTATTTGTGCAAT 232  
 Db 154818 TAATAATAAATAATATTATTCCTTAATAAATGACACATTTTAAATTTTACATTT 154759

QY 233 TTTTATTGTTGTACTGATTTACAGTAGTATTACATCATCAGTCACATACATCATCA 292  
- - - - -  
Db 154758 GTAATACAATATATTTTTGCATATATCATGAATATAATATATATATATATATATATA 154699

**Qy**    293 CATATATCACATATATTTTCTTAANTAAATAAAGGACTAAAATTCTCAACAC 352  
         |||||     |    |    |    |    |    |    |    |    |    |    |    |    |  
**Dd** 154698 TATATATATATATATATATAAATATATATATAATGTAGTATTTTTATATATATATAT 154630  
         |||||     |    |    |    |    |    |    |    |    |    |    |    |

QY 353 CAACGACATGTGAAGTTTTCTCCACAACACTTTTACCATTCTACATGTCCTATTTCGAA 412  
+ + + + +  
Db 154638 ATTATATATTCTGATAGATCTTCATATTAATGCCATATCATGAAAATAATGCATCAATGG- - - -  
+ + + + +

[illegible][illegible]

LOCUS	ATT25K17	8904 bp	DNA	linear	PLN 27-AUG-1999
DEFINITION	Arabidopsis thaliana	DNA	chromosome 4	BAC clone	(ESSA project).
ACCESSION	AF049171				

AL049171.1 GI:4539415  
AL049171

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis

1 (bases 1 to 89904)  
Bevan, M., Koetter, P., Hempel, S., Entian, K.-D., Bancroft, I.,  
Mewes, H.W., Mayer, K.F.X. and Schueller, C.  
Unpublished

2 (bases 1 to 89904)  
EU Arabidopsis sequencing, project.  
Direct Submission  
Submitted (12-MAR-1999), WTPS at the NCBI  
JOURNAL

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuellem@biochem.mpg.de, mayer@mps.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, Tebbs Institute, Cambridge University, Cambridge CB2 3RQ, UK

COMMENT

FEATURES

source	Location/Qualifiers 1. .89904 /organism="Arabidopsis thaliana"
viewed at:	<a href="http://websvr.mips.biochem.mpg.de/proj/thal/">http://websvr.mips.biochem.mpg.de/proj/thal/</a> .

```

/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"

misc_feature
1..11070

```

exon  
/note="overlap to BAC ; please refer to this entry for  
analysis and annotation"  
19417..19605  
/gene="m3c5v17\_10"

```

/9gene= 123K17.10"
/number=1
19417..21056
/gene="T25K17.10"
cns
424210127 1860

```

```

cds
join(19417. .13605,19706. .20001,20198. .21056)
/gene="T25K17.10"
/EC_number="4.4.1.14"

```

/note="strong similarity to ACC synthase , Malus domestica, U73816  
Contains Aminotransferases class-I pyridoxal-phosphate attachment site, Aa-Transfer Class\_1 [SLSKDLGLPGFRVG]  
contains EST gb:N65643, AA650853, Z33841"  
/codon\_start=1  
/product="1-aminocyclopropane-1-carboxylate synthase-like protein"  
/protein\_id="CAB38949.1"  
/db\_xref="GI:4539416"  
/translation="MGLPLMMERSNNNNVELSRVAVSDTHGEDSPYFAGWKAYDENP  
YDSEHNPNGVQIOMGLAGNOVSFDPLLETYLLEKNPEGSMMGSKGAPGRENALPDQYHG  
LKTFRAMASFMQIRGSKAFDPDRIVLTAGATAANELLFTIADPNDAALLVPTPT  
PGPDRDLRWETGQIVPIVHCDSNNHQITEALSAAYQTARDANIRVRGVLLITNPSP  
LGATVQKKVLEDDLDFCFCKRNKIHLSDEIYSGVSFHASETSVAEIVENIDDDYSVKR  
THVYISUKDGLGPGFRVGTIYSYNDNVNVARRMSSFTLVSSQTHMLASMLSDSEF  
TEKYINRERLRRRYDITVEGLKKAGIECLGNAGLFCWMNLGFLLEKTKDGEQLQ  
WDVILKELNLINISPGSSSCHSEVGFWRVCFANMSENTLEIALAKRIHEFMWDRRRR"  
19606..19705  
/gene="T25K17.10"  
/number=1  
19706..20001  
/gene="T25K17.10"  
/number=2  
20002..20197  
/gene="T25K17.10"  
/number=2  
20198..21056  
/gene="T25K17.10"  
/number=3  
26480..27228  
/gene="T25K17.20"  
join(26480..26661,27042..27228)  
/gene="T25K17.20"  
/note="strong similarity to predicted protein, Arabidopsis thaliana, AC005169"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB38950.1"  
/db\_xref="GI:4539417"  
/translation="WASKILQLSKACEASKRVSKHGTYYQLQDKKNMYTOEPATI  
EKNELSKQLLTTRASTIPGRSESPFKEVHNHVKGLWKNRADLUKVEDAGTAALFLGLEFC  
AWYCAEIVGRGFTTGYTP"  
26480..26661  
/gene="T25K17.20"  
/number=1  
26662..27041  
/gene="T25K17.20"  
/number=1  
27042..27228  
/gene="T25K17.20"  
/number=2  
28289..28369  
/gene="T25K17.30"  
/number=1  
28289..29256  
/gene="T25K17.30"  
join(28289..28369,28417..28641,28748..28979,28966..29256)  
/gene="T25K17.30"  
/note="strong similarity to caffeoyl-CoA  
O-methyltransferase, Pinus taeda, AF036095"  
/codon\_start=1  
/product="caffeoyl-CoA O-methyltransferase-like protein"  
/protein\_id="CAB38951.1"  
/db\_xref="GI:4539418"  
/translation="MAKDEAKGLLKSLEYKVCPIYLSTLYILETSVYPREPVLRE  
LVNTIHNHPQAGMATPDAGOLMGMLINLVNARTIEVGTGTSLLTALTLPEDGR  
LIAIDMNRSEIGLPIVKKAGVBEHKIDFKESEALPDALDELNNKVNVEGGDFAFVDA  
DKLWNYHERLIRLIVKGVLLIYDNTLWGGVSAPEDSSTPEIRVKKATLELNKLK  
SADRQVQLSQALGDGGITICRRLL"  
28370..28416  
/gene="T25K17.30"  
/number=1



```

/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
6687..8444
/gene="AT4g25770"
complement(6687..6833)
/gene="AT4g25770"
/number=1
complement(join(6687..6833,6920..7036,7083..7163,
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))
/gene="AT4g25770"
complement(join(6687..6833,6920..7036,7083..7163,
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))
/gene="AT4g25770"
/number=1
/translation="MTASFRRFOISIPDLIPSPRLPPSQNFRRLYLQFPAQPKN
LCOEMWNRIGICCFANRKRBLTTEIDNGDEDFDAVMESAERPDHLVVMVNGIV
GSRADWKYAAEQVKKFPDPKLVHRACFLTYRYCGSESATLTFDGVKMGRLANE
VFGCIYVSVLVGVRHRSGLKKISFVAHSLGLVARYAIGKLYEQGVEDSLDSPSK
EKARGGEIAGLEPMNFITPATPLGSRGHRFPILCGPLFLERTASQTAHLAAGRTG
KHLFLIDMLIISYDLNRSALNAFKRRVAVANRLSMVEDYFYHLQEFALTNEKAMVGW
RTSSIRRPNELRPNLLATDPNPHIVYVYVERGNVDNGSCQSTSTVVTQDTEGFLAL
FLN"
complement(6834..6919)
/gene="AT4g25770"
/number=1
complement(6920..7036)
/gene="AT4g25770"
/number=2
complement(7037..7082)
/gene="AT4g25770"
/number=2
complement(7083..7163)
/gene="AT4g25770"
/number=3
complement(7164..7277)
/gene="AT4g25770"
/number=3
complement(7278..7370)
/gene="AT4g25770"
/number=4
complement(7371..7493)
/gene="AT4g25770"
/number=4
complement(7494..7733)
/gene="AT4g25770"
/number=5
complement(7734..7781)
/gene="AT4g25770"
/number=5
complement(7782..7882)
/gene="AT4g25770"
/number=6
complement(7883..7944)
/gene="AT4g25770"
/number=6
complement(7945..8042)
/gene="AT4g25770"
/number=7
complement(8043..8133)
/gene="AT4g25770"
/number=7
complement(8134..8444)
/gene="AT4g25770"
/number=8
9569..10141
/gene="AT4g25780"
/number=1
/translation="MSSSVLRVLLLGALNVVVSLSITNSLTIKSATLGQVFRICKNL
CPGCDHDSIQFLFRHNLVRAAREPPLINDRRLONYAQWANQRGDCALRISVNGE
FNLGENIYSGANNSPADAVVAMASEKRFYHGSNTCDAGOMCGHYTIVWKSTRRY
9569..10141
/gene="AT4g25780"
/number=1
9569..10141
/gene="AT4g25780"
complement(10710..11074)
/gene="AT4g25790"
/number=1
10710..11752
/gene="AT4g25790"
complement(join(10710..11074,11485..11752))
/gene="AT4g25790"
complement(join(10710..11074,11485..11752))
/gene="AT4g25790"
/number=1
/translation="MATTSTQAIVFAIAELVIAOAVHADYRHRPQRPITPSPPPY
VAKPQLPTPSPKILYQPTTYQPTTGFQOFLDPHTVRGGLPLPLVMDVKIAS
YATWANRRYDCSLTSHSTGSGSDFTSTFAVSWTVEAKSYNHTMTCF
GDGCMGHYQIIVWRETRRLGCGAVVFCNYPGNYVYGEKPY"
complement(11075..11484)
/gene="AT4g25790"
/number=1
complement(11485..11752)
/gene="AT4g25790"
/number=2
13985..16227
/gene="AT4g25800"
join(13985..14010,14239..14669,14761..14921,15016..15219,
15390..15595,15645..16227)
/gene="AT4g25800"
/number=1
/translation="MNOFLVRLSLPLFTGGRVGEQGATIHVVLLDANTGRPVTVG
PEASIKLEVYVLGDFNNEDEDTQEFESVHVYKREGRKPLTGDLFVYLKRGVGT
LGEIVFDNSWIRSRFLGLRVPSCGYCIRIREAKTEAFSVYKDHGELYKKHYPP
ALNDEVWRLEKIGDGAFHKRLTAAGIVTVGFLRQLVSTDKLRAILGSGMKNMMD
LLVEHAKTCLVSLGKLYIYTEDSRVGVFNVIYELSLITEDYLSADSLSESQKVV
VDGLVKRAYENNVQVEGESLLNQLPERLDISQTDPTVTAALSYDSQFPEPA
IEGNIQISPLYSTIFTAGYNQTLTALPHNPQAQDFVPOQDFIGLIQQTQTNIE
NENYTRVLGPPQSTGGYQDIKKSADQENLPEDWTNLSEDDFFSEEEIQTQSHDL
LANEDMQQLFSMGGGEGDGFTFPSFMQNPMMQGYDEEGRGSRGKAVVGLVKKAA
MRWGGFFRRKAAERAAQIVELHNDNGK"
13985..14010
/gene="AT4g25800"

```

```

/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
6687..8444
/gene="AT4g25770"
complement(6687..6833)
/gene="AT4g25770"
/number=1
complement(join(6687..6833,6920..7036,7083..7163,
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))
/gene="AT4g25770"
complement(join(6687..6833,6920..7036,7083..7163,
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))
/gene="AT4g25770"
/number=1
/translation="MTASFRRFOISIPDLIPSPRLPPSQNFRRLYLQFPAQPKN
LCOEMWNRIGICCFANRKRBLTTEIDNGDEDFDAVMESAERPDHLVVMVNGIV
GSRADWKYAAEQVKKFPDPKLVHRACFLTYRYCGSESATLTFDGVKMGRLANE
VFGCIYVSVLVGVRHRSGLKKISFVAHSLGLVARYAIGKLYEQGVEDSLDSPSK
EKARGGEIAGLEPMNFITPATPLGSRGHRFPILCGPLFLERTASQTAHLAAGRTG
KHLFLIDMLIISYDLNRSALNAFKRRVAVANRLSMVEDYFYHLQEFALTNEKAMVGW
RTSSIRRPNELRPNLLATDPNPHIVYVYVERGNVDNGSCQSTSTVVTQDTEGFLAL
FLN"
complement(6834..6919)
/gene="AT4g25770"
/number=1
complement(6920..7036)
/gene="AT4g25770"
/number=2
complement(7037..7082)
/gene="AT4g25770"
/number=2
complement(7083..7163)
/gene="AT4g25770"
/number=3
complement(7164..7277)
/gene="AT4g25770"
/number=3
complement(7278..7370)
/gene="AT4g25770"
/number=4
complement(7371..7493)
/gene="AT4g25770"
/number=4
complement(7494..7733)
/gene="AT4g25770"
/number=5
complement(7734..7781)
/gene="AT4g25770"
/number=5
complement(7782..7882)
/gene="AT4g25770"
/number=6
complement(7883..7944)
/gene="AT4g25770"
/number=6
complement(7945..8042)
/gene="AT4g25770"
/number=7
complement(8043..8133)
/gene="AT4g25770"
/number=7
complement(8134..8444)
/gene="AT4g25770"
/number=8
9569..10141
/gene="AT4g25780"
/number=1
/translation="MSSSVLRVLLLGALNVVVSLSITNSLTIKSATLGQVFRICKNL
CPGCDHDSIQFLFRHNLVRAAREPPLINDRRLONYAQWANQRGDCALRISVNGE
FNLGENIYSGANNSPADAVVAMASEKRFYHGSNTCDAGOMCGHYTIVWKSTRRY
9569..10141
/gene="AT4g25780"
/number=1
9569..10141
/gene="AT4g25780"
complement(10710..11074)
/gene="AT4g25790"
/number=1
10710..11752
/gene="AT4g25790"
complement(join(10710..11074,11485..11752))
/gene="AT4g25790"
complement(join(10710..11074,11485..11752))
/gene="AT4g25790"
/number=1
/translation="MATTSTQAIVFAIAELVIAOAVHADYRHRPQRPITPSPPPY
VAKPQLPTPSPKILYQPTTYQPTTGFQOFLDPHTVRGGLPLPLVMDVKIAS
YATWANRRYDCSLTSHSTGSGSDFTSTFAVSWTVEAKSYNHTMTCF
GDGCMGHYQIIVWRETRRLGCGAVVFCNYPGNYVYGEKPY"
complement(11075..11484)
/gene="AT4g25790"
/number=1
complement(11485..11752)
/gene="AT4g25790"
/number=2
13985..16227
/gene="AT4g25800"
join(13985..14010,14239..14669,14761..14921,15016..15219,
15390..15595,15645..16227)
/gene="AT4g25800"
/number=1
/translation="MNOFLVRLSLPLFTGGRVGEQGATIHVVLLDANTGRPVTVG
PEASIKLEVYVLGDFNNEDEDTQEFESVHVYKREGRKPLTGDLFVYLKRGVGT
LGEIVFDNSWIRSRFLGLRVPSCGYCIRIREAKTEAFSVYKDHGELYKKHYPP
ALNDEVWRLEKIGDGAFHKRLTAAGIVTVGFLRQLVSTDKLRAILGSGMKNMMD
LLVEHAKTCLVSLGKLYIYTEDSRVGVFNVIYELSLITEDYLSADSLSESQKVV
VDGLVKRAYENNVQVEGESLLNQLPERLDISQTDPTVTAALSYDSQFPEPA
IEGNIQISPLYSTIFTAGYNQTLTALPHNPQAQDFVPOQDFIGLIQQTQTNIE
NENYTRVLGPPQSTGGYQDIKKSADQENLPEDWTNLSEDDFFSEEEIQTQSHDL
LANEDMQQLFSMGGGEGDGFTFPSFMQNPMMQGYDEEGRGSRGKAVVGLVKKAA
MRWGGFFRRKAAERAAQIVELHNDNGK"
13985..14010
/gene="AT4g25800"

```

overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tldb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<ftp://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: [at@tigr.org](mailto:at@tigr.org).

FEATURES	source	Location/Qualifiers	
		1..115522	
		/organism="Arabidopsis thaliana"	
		/cultivar="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="II"	
misc_feature		complement(<1..1408)	
repeat_region		/note="Sequence from clone F26H6"	
		515..628	
mrna		/rpt_family="(TAAAA)n"	
		join(<940..1083,1286..1597,1804..2328,2426..2653)	
gene		/gene="At2g15460"	
		<940..>2653	
		/gene="At2g15460"	
		/note="F26H6.2"	
CDS		join(940..1083,1286..1597,1804..2328,2426..2653)	
		/gene="At2g15460"	
		/codon_start=1	
		/product="putative polygalacturonase"	
		/protein_id="AA017390.2"	
		/db_xref="GI:6598543"	
		/translation="MACISFAFALCFLFFIVVASSTVRPKVFNVRGAKPDGKT DNAAFTSWSRACKRISGSKIYVPGTFLGVGEVGPCKNPDEFIDTGLIAPN PSDIKQDTWINFRIYNNLSISGGSTLDGQKQWPHNDHTNPNCKLMTGMFAFVN NSNFKDTLSNKMGNHFNFTSVHFNITGVITAPGDSPTDGIKMGSCSNIGTSDTN LGTGDCDIAILSGTTNLINSNVCGPGHGISVSGSKSKDEKDKVDLIVRDFNGTS DGRIKTWESSAKILYSNFEVNIOMIDYKPINIDQKYPHPPCPCHERRKGSHVOI QNKLKNIYGTSTSKNVAVNLQCKSIFPCKNVELIDINIKQNGYKDGSTSVCEVNDGF ARGKMFPRIV"	
misc_feature		1409..115321	
repeat_region		/note="Sequence from clone F9013"	
		complement(2357..2378)	
repeat_region		/rpt_family="(TAAAA)n"	
		3911..4024	
mrna		/rpt_family="(TAAAA)n"	
		join(<4336..4479,4682..4993,5202..5726,5830..>6057)	

intron	/number=1		Query Match	4.1%; Score 54; DB 8; Length 196286;	Best Local Similarity 56.6%; Pred. No. 0.11;	Matches 120; Conservative 0; Mismatches 90; Indels 2; Gaps 1;
	14011..14238					
	/gene="AT4g25800"					
	/number=1					
exon	14239..14669		Mb 169602	ATTAAATTTAATAATAATAATTACTAAATAATCATTTATTGGAATTTTGGTTATTATTTT 169661		
	/gene="AT4g25800"					
	/number=2					
	14670..14760					
intron	/gene="AT4g25800"		Mb 319	AAATTTAAACTAAAATGACTAAATTTCTAAACCAACGACGATGTAATGTTTCTCCAA 378		
	/number=2					
	14761..14921					
	/gene="AT4g25800"					
exon	/number=3		Mb 169662	CATTTAAATTTAAAGTCTGTATAAATTTAAATAATTTATATATATAAATTTTCTTCAA 169721		
	14922..15015					
	/gene="AT4g25800"					
	/number=3					
intron	/number=3		Mb 379	CAACTTTTACCTTCTTCACATGTTCTATTTTCGAATTTCACTCTATAAACACATAGTCTA 438		
	14922..15015					
	/gene="AT4g25800"					
	/number=3					
intron	/number=3		Mb 169722	AACATTTATTATTATTATA--GTAAATTTTCAATTATATACAAAGTTAAATAAGTTT 169779		
	14922..15015					
	/gene="AT4g25800"					
	/number=3					
intron	/number=3		Mb 439	CAATGAAACACAGTGCTTTTGACGACTATATA 470		
	14922..15015					
	/gene="AT4g25800"					
	/number=3					
exon	/number=3		Mb 169780	CACTTTTAAACATTTATTTTAACTAAATTTTAA 169811		
	14922..15015					
	/gene="AT4g25800"					
	/number=3					

gene  
/gene="At2g15470"  
<4336. .>6057  
/gene="At2g15470"  
/note="F9013.2"  
join(4336. .4479,4682. .4993,5202. .5726,5830. .6057)  
/gene="At2g15470"  
/codon\_start=1  
/product="putative polygalacturonase"  
/protein\_id="AADI7391.1"  
/db\_xref="GI:4335713"  
/translation="MACTSFAPKALCFLFFIVASRSVTRPKVFNQVRHGAQPKGT  
DNANFTSIRACRKSIGSSKIYVPGTFYLGGVFGLGNPCNPIEFIDGTLAPAN  
PSDIKDTWIRFYINLISIGSTLDGQKQSWHNDCTNPNCKLAMTWGFARFN  
NSNIGKSLNKGHFNFYSVHHFNITGVTTAPGDSPTDGIKMGSCSNQISPTN  
IGTGDDCIAILSGTNLNSVNCGPGHGISVSGLGKNDKDKADLLIVRDVIFNFTS  
DGIKRTWESSAKILYSNFYENIQMIDVGPINIDQKYPCHPCEHERKSHVQION  
LKLNIYTSKNKVAWNLQCSKIFPCKNVELIDINIKQNGVKDGSSTVCENVDFGAR  
GMFPOHCLN"  
complement(5755. .5776)  
/rpt\_family="(TAAA)n"  
7220. .7281  
/rpt\_family="(TA)n"  
complement(7698. .7802)  
/rpt\_family="(TA)n"  
complement(7735. .7838)  
/rpt\_family="(TAAA)n"  
complement(8026. .8142)  
/rpt\_family="(TAA)n"  
complement(8080. .8183)  
/rpt\_family="(GAAA)n"  
complement(8739. .8780)  
/rpt\_family="(CAAAA)n"  
join(<8823. .9813,9995. .>10458)  
/gene="At2g15480"  
<8823. .>10458  
/note="F9013.3"  
/gene="At2g15480"  
join(8823. .9813,9995. .10458)  
/gene="At2g15480"  
/codon\_start=1  
/product="putative glucosyltransferase"  
/protein\_id="AADI7392.1"  
/db\_xref="GI:4335714"  
/translation="MNRYSERHILFFPFMAQGHMIPILMAKLFRRGAKSTLLTT  
PINAKFEPIPAKFNQNDLEIGIKIENPCVELGLPEGCEADNADINSYQSDSDGL  
FLKFEFTYMKQLESFIETTKPSALVADMEFPWATESAEKLGVPRLVHGFSPFSL  
CCSNHRIHKPKKATSTPFPVIGLPGDIVITEDQANVAKEETPMKPFKEVRESE  
TNSFGVLVSFYELAYSADFYRSFAKRAWHIGPLSLNRELGERARRGKKNIDEQ  
ECLKWLDSKTPGSVVYLSFGSGTFTNDQLLEIAFLEGSGQSFIVWVRKNENQDGE  
EWLPGFKERTTGKGLIPGWAPOVLIDHKAIGGFVTHCGWNSATEGLAAGLPVMTW  
PWAEQFYNEKLLTKVLRIGVNVGATVELYKCKLISRAQVEKAVREVIIGERAEERL  
WAKLGEKMAKAAVEEGSSYNDVNKFMEELENGRK"  
join(<11828. .12737,12932. .>13404)  
/gene="At2g15490"  
<11828. .>13404  
/note="F9013.4"  
join(11828. .12737,12932. .13404)  
/gene="At2g15490"  
/codon\_start=1  
/product="putative glucosyltransferase"  
/protein\_id="AADI7393.1"  
/db\_xref="GI:4335715"  
/translation="MAKLFARGAKSTLLTTPINAKILEKPIEAFKVPQNDLEIGIKI  
LNFPCVELPGECEGNRFINSYQSDSDFLFLKFLFYMKQLESFIETTKPSAL  
LVADFPWATESAEKLGVPRLVHGFSPFALCCSNHRIHKPKKVASSTPFIPLG  
PGDIVITEDQANVEETPMKPFKEVRESESSFGVLVNSFYELSSVDFYRSFA  
KWAHIGPLSLNRGTAERKAGKKNIDOECLKWLDSKTPGSVYLSFGSGTGPLN  
EOLLEIAFLEGSGONFIWVSKNENQVTGENEDWLPKGFERNKKGKLLIRGWRPQ  
VILLDHKAIGGFVTHCGWNSLLEGTAAGLPMTWPMGAEQFYNEKLLTKVLRIGVNVG  
ATELVKKGKLLISRAQVEKAVREVIIGERAEERLRLKELGEKMAKAAVEEGSSYNDVN  
KFMEELENGRK"  
join(<13713. .13796,13945. .14093,14268. .14303,14375. .14413,

gene  
14667. .>14757)  
/gene="At2g15500"  
<13713. .>14757  
/gene="At2g15500"  
/note="F9013.5"  
join(13713. .13796,13945. .14093,14268. .14303,14375. .14413,  
14667. .14757)  
/gene="At2g15500"  
/codon\_start=1  
/product="putative poly(A)-binding protein"  
/protein\_id="AADI7394.1"  
/db\_xref="GI:4335716"  
/translation="MSKLGMLLNDKQVYVGPFLRKGKGLLAVESLKGNTFDKQWV  
VRAOKKSREMELKVPYEQSLKEIADKFQSYNFRSGFVAFSTQNTVTSIGICKQVID  
ESVAMPSPVGRPMPIYPLSGPGIRKSMFYG"  
complement(join(<16628. .18427,18538. .18661,18914. .19100,  
19343. .19660,19695. .>20682))  
/gene="At2g15510"  
complement(<16628. .>20682)  
/gene="At2g15510"  
/note="F9013.6"  
Query Match 4.1%; Score 53.8; DB 8; Length 115522;  
Best Local Similarity 53.3%; Pred. No. 0.1;  
Matches 136; Conservative 0; Mismatches 117; Indels 2; Gaps 1;  
QY 226 TGCATATTTTATTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285  
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 8079 TGATTATTGTAAGTAGTAATCAAAAGGTGAATTAATAGTATTTAACAATAATAATAGTAA 8020  
QY 286 ATCATACATATATACAAATATTTTCTAAATTAATTAATAAACTAAAAATGACTAAATTT 345  
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 8019 AAATCATTTTATGATTTTGGTTTATTAATTCATTTAAATTAAGTCTGCATAAAT 7960  
QY 346 CTAACACCAACGACATGTAAATGTTTCTCCAAACAACTTACCTATTCTACATGTTTCTA 405  
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 7959 TTAATAATAATATATATATAAATTTCTCTTCAAAACATTTTATTTATTTATAAAAAA 7900  
QY 406 TTTGCAATTTCACTCTATAACAACATAGTCTACAATGGAAACAGTCGTTTGTACCACT 465  
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 7899 TT--AAATATATATACAAATTTAAATAAGTTTCACTTTAAACACATTTATTTATACAAAT 7842  
QY 466 ATATACGCGATGTGT 480  
Db 7841 TTAATAATAATTTAT 7827  
RESULT 14  
AC105576/c 214033 bp DNA linear HTG 09-JAN-2002  
LOCUS Rattus norvegicus clone CH230-236F15, WORKING DRAFT SEQUENCE, 50  
unordered pieces.  
AC105576 GI:18092798  
AC105576.1 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Norway rat.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 214033)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Boucek,J.,  
Bowle,S., Brileva,M., Brown,M., Brown,E., Brown,N.P., Buhay,C.,  
Burch,P., Burrett,C., Burrell,K.L., Byrd,N.C., Cartron,F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Drepper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Correll, J. H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C.,  
 Lewis, B., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louisgied, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
 Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
 Ruiz, S., Savory, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,  
 Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, K., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 214033)  
 Worley, K. C.  
 Direct Submission  
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 -----  
 Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 -----  
 Project Information  
 Center project name: GNFN  
 Center clone name: CH230-236F15  
 -----  
 Summary Statistics  
 Assembly program: Phrap; version 0.990329 First call to  
 findPhrapList  
 Consensus quality: 185044 bases at least Q40  
 Consensus quality: 190296 bases at least Q30  
 Consensus quality: 194221 bases at least Q20  
 Estimated insert size: 190368; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Cenbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Cenbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 50 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 16332: contig of 16332 bp in length  
 \* 16333: gap of unknown length  
 \* 16433: contig of 10459 bp in length  
 \* 26891: gap of unknown length  
 \* 26892: contig of 12123 bp in length  
 \* 39114: contig of 8492 bp in length  
 \* 39115: gap of unknown length  
 \* 39215: contig of 8492 bp in length  
 \* 47707: gap of unknown length  
 \* 47807: contig of 7099 bp in length  
 \* 54905: contig of 7099 bp in length  
 \* 54906: gap of unknown length

55006: contig of 6725 bp in length  
 61731: gap of unknown length  
 61830: contig of 7092 bp in length  
 68922: gap of unknown length  
 69022: contig of 4933 bp in length  
 69023: gap of unknown length  
 73955: contig of 5884 bp in length  
 74055: gap of unknown length  
 79739: contig of 5884 bp in length  
 79839: gap of unknown length  
 79840: contig of 7364 bp in length  
 87203: gap of unknown length  
 87204: contig of 8413 bp in length  
 95716: gap of unknown length  
 95717: contig of 5404 bp in length  
 95817: gap of unknown length  
 101221: contig of 5404 bp in length  
 101320: gap of unknown length  
 106991: contig of 5671 bp in length  
 107091: gap of unknown length  
 107092: contig of 4093 bp in length  
 111184: gap of unknown length  
 111284: gap of unknown length  
 111285: contig of 5718 bp in length  
 117002: gap of unknown length  
 117003: contig of 5617 bp in length  
 122719: gap of unknown length  
 122720: contig of 4857 bp in length  
 122820: gap of unknown length  
 127676: contig of 4857 bp in length  
 127677: gap of unknown length  
 132248: contig of 4472 bp in length  
 132348: gap of unknown length  
 13419: contig of 5071 bp in length  
 137519: gap of unknown length  
 137520: contig of 3000 bp in length  
 140519: gap of unknown length  
 140520: contig of 4994 bp in length  
 140620: gap of unknown length  
 145613: contig of 4994 bp in length  
 145614: gap of unknown length  
 145713: contig of 5030 bp in length  
 150743: gap of unknown length  
 150843: contig of 2386 bp in length  
 153229: gap of unknown length  
 153230: contig of 2376 bp in length  
 15330: gap of unknown length  
 15330: contig of 2376 bp in length  
 155705: gap of unknown length  
 155706: contig of 3836 bp in length  
 155805: gap of unknown length  
 159641: gap of unknown length  
 159642: contig of 2196 bp in length  
 161937: gap of unknown length  
 162037: contig of 4099 bp in length  
 162038: gap of unknown length  
 166137: contig of 4099 bp in length  
 166236: gap of unknown length  
 166237: contig of 3418 bp in length  
 169654: gap of unknown length  
 169755: contig of 3455 bp in length  
 173209: gap of unknown length  
 173210: contig of 1563 bp in length  
 174872: gap of unknown length  
 174873: contig of 3398 bp in length  
 178371: gap of unknown length  
 178470: contig of 3296 bp in length  
 181766: gap of unknown length  
 181767: contig of 2311 bp in length  
 181867: gap of unknown length  
 184177: contig of 1976 bp in length  
 184278: gap of unknown length  
 186253: contig of 2536 bp in length  
 186354: gap of unknown length  
 188899: gap of unknown length  
 188900: contig of 1741 bp in length  
 190730: gap of unknown length  
 190731: contig of 1407 bp in length  
 190831: gap of unknown length  
 192238: contig of 2129 bp in length  
 192338: gap of unknown length  
 194466: gap of unknown length  
 194567: contig of 1658 bp in length  
 196224: gap of unknown length  
 196325: contig of 1470 bp in length  
 196325: gap of unknown length  
 197794: contig of 1477 bp in length  
 197795: gap of unknown length  
 197895: contig of 1477 bp in length  
 199371: gap of unknown length  
 199372: contig of 2793 bp in length  
 199471: gap of unknown length  
 202264: contig of 2793 bp in length

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

\* 202265 202364: gap of unknown length  
 \* 202365 204373: contig of 2009 bp in length  
 \* 204374 204473: gap of unknown length  
 \* 204474 205726: contig of 1253 bp in length  
 \* 205727 205826: gap of unknown length  
 \* 205827 207670: contig of 1844 bp in length  
 \* 207671 207770: gap of unknown length  
 \* 207771 208920: contig of 1050 bp in length  
 \* 208921 208920: gap of unknown length  
 \* 208921 210151: contig of 1231 bp in length  
 \* 210152 210251: gap of unknown length  
 \* 210252 211652: contig of 1401 bp in length  
 \* 211653 211752: gap of unknown length  
 \* 211753 212869: contig of 1117 bp in length  
 \* 212870 212969: gap of unknown length  
 \* 212970 214033: contig of 1064 bp in length.

# FEATURES

Location/Qualifiers  
 1..214033  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-236F15"

BASE COUNT 63123 a 41277 c 42183 g 52289 t 5161 others  
 ORIGIN

Query Match 4.1%; Score 53.6; DB 2; Length 214033;  
 Best Local Similarity 50.8%; Pred. No. 0.13;  
 Matches 128; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
 QY 815 CCGTGGCCACCTCCAGTCCACGGCAGCATCCCTCTGAGCGGCTCACCAGCGATGCC 874  
 DB 140327 CCCCAGACCCCGCTCCCGCCGCCGCCCTCTGCTCTGACCCGCCGCCGCCGCC 140268  
 QY 875 GTGTCGGGAGCGGACCAACCAACCCCAACCACTCAGCAAAACCCCGCCGCCGTCGC 934  
 DB 140267 TCCCTTCCCGCCCGCTCCCGCCGCCGCCCGCCGCCCGCTCCGCCGCCGCCGCC 140208  
 QY 935 CCGTGGTCCGCGTCCGGCAACGAGCGCGCGCGTGTGAGTCCCTGGACACCGCA 994  
 DB 140207 CCGCTGCTCCCGCCGCCGCCGCCCGCTCCGCGGCGTCCGCCGCCGCCGCCGCC 140148  
 QY 995 CACCTGTGCGGCCCTTTCTTTATTCATCCCGAATCTCATCTGCCCGCCAGCGCGACTGC 1054  
 DB 140147 TCCCTTCCCGCCCGCTCCCGCCGCCCGCTCCCGCCCGCTCCCGCTCCCGCTCC 140088  
 QY 1055 GCTGGCGCGCCCG 1066  
 DB 140087 CCGCGCGCCCGC 140076

## RESULT 15

AF254090  
 LOCUS  
 DEFINITION Rhopalosiphum maidis 614 bp DNA linear INV 06-DEC-2000  
 partial cds; mitochondrial gene for mitochondrial product.  
 AF254090  
 VERSION  
 AF254090.1 GI:11559553  
 SOURCE  
 corn leaf aphid.  
 ORGANISM  
 Mitochondrion Rhopalosiphum maidis  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;  
 Aphidiformes; Aphidoidea; Aphididae; Aphidini; Rhopalosiphum.  
 1 (bases 1 to 614)  
 Chen, Y., Giles, K.L., Payton, M.E. and Greenstone, M.H.  
 Identifying key cereal aphid predators by molecular gut analysis  
 Mol. Ecol. 9 (11), 1887-1898 (2000)  
 MEDLINE  
 20546168  
 PUBMED  
 11091324

## REFERENCE

AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 2 (bases 1 to 614)  
 Chen, Y., Giles, K.L., Payton, M.E. and Greenstone, M.H.  
 Direct Submission  
 Submitted (10-APR-2000) Entomology and Plant Pathology, Oklahoma  
 State University, 127/110 Noble Research Center, Stillwater, OK

74078-3033, USA

FEATURES  
 source

Location/Qualifiers  
 1..614  
 /organism="Rhopalosiphum maidis"  
 /organelle="mitochondrion"  
 /db\_xref="taxon:43146"

gene  
 <1..>614

CDS  
 /gene="COII"  
 <1..>614  
 /gene="COII"  
 /codon\_start=1  
 /transl\_table=5  
 /product="cytochrome oxidase subunit II"  
 /protein\_id="AAG38002.1"  
 /db\_xref="GI:11559554"  
 /translation="LIFFHDHTFIILIMTITTYMMIFIMKKNKINIKISENQMLEF  
 IWTTPPIILIFIAPLSLHLLYLMDEIKCPILTIKIFGHQWFWHSYSEFSNIEFESY  
 MNELNKENFRLLIEVDNKTIIPEKFNIRLLISSDDVIHWSWTIPSLAIKIDAIPGRMNO  
 INLFMRPGMYFCQCEIGINHSFPIQIESINLNKFIYIKN"

BASE COUNT 247 a 69 c 48 g 250 t

ORIGIN

Query Match 4.1%; Score 53.4; DB 3; Length 614;  
 Best Local Similarity 56.3%; Pred. No. 0.038;  
 Matches 120; Conservative 0; Mismatches 91; Indels 2; Gaps 1;

QY 235 TTATTGTTTACTGATTATACGAGTAGTATATACATACATACATACATACATACATAC 294  
 DB 1 TTAATTTTTTTCATGATCATAAATTTTATTATTATTATTATTATAACAACAATTA 60

QY 295 TATATCACATATTTTCTAAATTAATTAATAAATAAATAAATAAATAAATAAATAA 354  
 DB 61 TATATAATA--ATTTTTATTATAAATAAATAAATAAATAAATAAATAAATAAAT 118

QY 355 ACACATTTGAATGTTTCTCCCAACACTTTACCTATTCTACATGTTCTATTTCGAATT 414  
 DB 119 AAATAATTGAATTATTGTAACACTACACTCCACTATTATTCTATTTTTCGTATAC 178

QY 415 TCACCTATATAACAACATAGCTCTACATGGAATGAAA 447  
 DB 179 CTTCTTACATCTATTATATCTAATAATAGATAA 211

Search completed: November 7, 2002, 12:14:29  
 Job time : 3553.14 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:29:53 ; Search time 166.564 Seconds  
(without alignments)  
13503.300 Million cell updates/sec

Title: US-09-905-558c-3

Perfect score: 1310

Sequence: 1 cccatgcgtcttctgtctac.....caagggaagtgatcgatg 1310

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-nbml/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235	17.9	12313	21	AAS35184
2	232.8	17.8	2126	24	AAS96571
3	124.6	9.5	1888	22	AAF81468
4	50.2	3.8	6327	22	AAS45444
5	50	3.8	10133	24	ABL32458
6	49.4	3.8	7334	24	ABL34124
7	49.2	3.8	11691	24	ABL34241
8	48.8	3.7	1337	20	AAZ17263
9	48	3.7	399	22	AAH93294
					Corn delta-12 desa
					Corn promoter sequ
					Corn promoter clon
					Chemically pretrea
					Human immune syste
					Human immune syste
					Human gene express
					Plasmodium falcipa

C 10	47.8	3.6	326	22	AAS56305	Human CDNA for an
C 11	47.8	3.6	73334	24	ABL34125	Human immune syste
C 12	47.4	3.6	6627	24	ABL33574	Human immune syste
C 13	47.2	3.6	6033	21	AA70152	Plasmodium falcipa
C 14	46.8	3.6	281	22	AA16482	Human breast cance
C 15	46.8	3.6	6107	24	AAS61315	Human gene regulat
C 16	46.6	3.6	20420	22	AAK73165	Human immune/haema
C 17	46.4	3.5	463	22	AAS56574	Human CDNA for an
C 18	46	3.5	6988	24	ABL34440	Human immune syste
C 19	45.8	3.5	2140	21	AAZ49411	Pea DRR206 protein
C 20	45.8	3.5	8991	24	AAS61295	Human gene regulat
C 21	45.6	3.5	3730	22	ABA19745	Human nervous syst
C 22	45.6	3.5	6699	24	AAS63321	Chemically pretrea
C 23	45.6	3.5	7049	24	ABL32157	Human immune syste
C 24	45.2	3.5	4257	19	AAV68520	The nucleotide seq
C 25	45.2	3.5	4257	19	AAV10362	Infected cell prot
C 26	45.2	3.5	12001	16	AAQ76213	HSV L/SST region.
C 27	45.2	3.5	13326	24	ABL33712	Human immune syste
C 28	45.2	3.5	17918	24	AAS61418	Human gene regulat
C 29	45	3.4	687	24	ABL34283	Human immune syste
C 30	45	3.4	883	22	AA226436	Human breast cance
C 31	45	3.4	883	22	AA226587	Human breast cance
C 32	45	3.4	883	22	AA226703	Human breast cance
C 33	45	3.4	883	22	AA226709	Human breast cance
C 34	45	3.4	883	22	AA226724	Human breast cance
C 35	45	3.4	883	22	AA226756	Human breast cance
C 36	45	3.4	883	22	AA226774	Human breast cance
C 37	45	3.4	883	22	AA226785	Human breast cance
C 38	45	3.4	6478	22	AAS45417	Chemically pretrea
C 39	45	3.4	6626	22	AAS46810	Tumour suppressor
C 40	45	3.4	10329	24	ABL34123	Human immune syste
C 41	45	3.4	16217	24	ABL32625	Human immune syste
C 42	44.8	3.4	259	22	AA225325	Human breast cance
C 43	44.8	3.4	11473	24	ABL33355	Human immune syste
C 44	44.8	3.4	17419	22	AAS45393	Chemically pretrea
C 45	44.8	3.4	17419	24	ABL33295	Human immune syste

ALIGNMENTS

RESULT 1	
AAZ35184	
ID	AAZ35184 standard; DNA; 12313 BP.
XX	
AC	AAZ35184;
XX	
DT	13-MAR-2000 (first entry)
XX	
DE	Corn delta-12 desaturase fad2-2 gene 5' untranslated region.
XX	
KW	Corn; maize; transgenic plant; lipid; food; feedstuff;
KW	vegetable oil; seed oil; oleic acid; fatty acid desaturase;
KW	delta-12 desaturase; fad2-2; ss.
XX	
OS	Zea mays.
XX	
FH	Key
FT	Intron
FT	Location/Qualifiers
FT	5651..12301
FT	/*tag= a
FT	5439..5444
FT	/*tag= b
XX	
PN	WO964579-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-US12884.
XX	
XX	11-JUN-1998; 98US-0088987.
PR	
PA	(DUPO ) DU PONT DE NEMOURS & CO E. I.
XX	

```
PI Shen JB;
XX WPI; 2000-097535/08.
XX New maize oleosin promoter, used for producing transgenic plants with
XX altered fatty acid composition of the oil, used e.g. in animal feeds
XX Example 2; Page 80-84; 108pp; English.
XX This is the nucleotide sequence of the 5' untranslated region of
XX the novel corn fad2-2 gene that codes for delta-12 desaturase (see
XX AAY32344). The invention generally relates to the preparation and use
XX of nucleic acid fragments comprising all, or substantially all, of a
XX corn oleosin promoter (see AA35165-77), a stearyl-ACP desaturase
XX (see AA35179-80) and a delta-12 desaturase, which can be used
XX individually or in combination to modify the lipid profile of corn.
XX Suppression of delta-12 desaturase expression can increase the
XX oleic acid content of the seed oil. Also claimed are seeds of
XX such plants; oil obtained from the grain of such plants; animal
XX feed, use of the oil in food, feed, and cooking oil or industrial
XX applications. The promoter region of fad2-2 may be used to
XX express a gene of interest in transgenic corn plants.
XX Sequence 12313 BP; 2845 A; 2969 C; 3170 G; 3329 T; 0 other;

Query Match 17.9%; Score 235; DB 21; Length 12313;
Best Local Similarity 82.5%; Pred. No. 3.3e-50;
Matches 293; Conservative 0; Mismatches 60; Indels 2; Gaps 2;

QY 5 TCGCTGCTTTGTCTACATCATCTTCTTCATCATCTCTCCCGGCGACGGTCTGCTGTT 64
DB 3444 TCGCGTGATCGTCCACAGCTTGCTGTCGCTTACCCTACCCAGTTGACCGTCTGCTGTT 3503
QY 65 CTTATTCA-GACTACCGTTCGAGTGACATGCGGTACATCTTTCTGATCGACTTTGT 123
DB 3504 CTTCTCCGGCGGACCGTTCGAGGACTGCACTGCGTACACCTTCTGCACGACTTCGT 3563
QY 124 ACGGTACATCGAATATACACGAGATGTCCTGTTGAATAGATCACTAATGCCTTAA 183
DB 3564 ACGACTACATCGAACAACACACGAGATGTCCTGTTGAATGGAGCCACTGGTGCCTTGA 3623
QY 184 GCATCGGTACTCCGTAGGTACATCTCTTCTTCTTTATTTGTGCATATTTTATTGTTG 243
DB 3624 GCATCGGTCCCTCCGCTGGGTACATCTGTTCTGTTATTTGTGCAT-GTTTCATGCTG 3682
QY 244 TTTACTGATTATACGAGTAGTTATACATACATGACATACATATACATATATACACA 303
DB 3683 TTTACTGCTTATGCGAGTAGTTATACACATGACATACATATGTCATACATATATACA 3742
QY 304 ATATTTTCTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 358
DB 3743 CTGATTTTCTGGATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3797

RESULT 2
AAS96571
ID AAS96571 standard; DNA; 2126 BP.
XX
AC AAS96571;
XX
XX 26-FEB-2002 (first entry)
XX
DE Corn promoter sequence #14.
XX
XX Corn; male reproductive tissue; plant regulatory sequence; Zea mays;
XX promoter; transcription regulation; operably linked gene; monocot;
XX dicot; wheat anther; plant fertility; insect tolerance;
XX pathogen tolerance; herbicide tolerance; ds.
XX
OS Zea mays.
XX
XX WO200183790-A2.
XX
```

```
PD 08-NOV-2001.
XX
XX 30-APR-2001; 2001WO-US13739.
XX
PR 01-MAY-2000; 2000US-201255P.
XX
XX (MONS ) MONSANTO TECHNOLOGY LLC.
XX
XX Conner TW, Dubois P, Malven M, Masucci JD;
XX WPI; 2002-055481/07.
XX
XX Novel promoters isolated from corn for controlling gene expression in
XX male reproductive tissues, such as anthers, tassels, and to regulate
XX transcription of target genes including genes for insect or pathogen
XX tolerance
XX
XX Claim 1; Page 115-116; 121pp; English.
XX
XX The present invention relates to the isolation of plant regulatory
XX sequences from the male reproductive tissues of corn (Zea mays). The
XX promoter sequences, fragments, regions or cis elements of the sequences,
XX are capable of regulating transcription of an operably linked DNA
XX sequence. The promoter sequences confer enhanced expression of operably
XX linked genes in monocot or dicot male reproductive tissues, such as
XX anthers, especially wheat anthers and is useful for regulating
XX transcription of a DNA sequence, by operably linking the DNA sequence
XX to the promoter. The promoter sequences are useful in plants to regulate
XX transcription of target genes including genes for control of fertility,
XX insect or pathogen tolerance and herbicide tolerance. They are also
XX useful as probes or primers in nucleic acid hybridisation experiments.
XX The promoter sequences can be used in hybridisation assays of other
XX plant tissues to identify closely related or homologous genes and
XX associated regulatory sequences. AAS96558-AAS96577 represent the
XX corn promoter sequences of the present invention.
XX
XX Sequence 2126 BP; 607 A; 419 C; 433 G; 567 T; 0 other;

Query Match 17.8%; Score 232.8; DB 24; Length 2126;
Best Local Similarity 83.0%; Pred. No. 5.6e-50;
Matches 289; Conservative 0; Mismatches 57; Indels 2; Gaps 2;

QY 5 TCGCTGCTTTGTCTACATCATCTTCTTCATCATCTCTCCCGGCGACGGTCTGCTGTT 64
DB 82 TCGCGTGATCGTCCACAGCTTGCTGTCGCTTCCCAAGTTGACCGTCTGCTGTT 141
QY 65 CTTATTCA-GACTACCGTTCGAGTGACATGCGGTACATCTTTCTGCATCGACTTTGT 123
DB 142 CTTCTCCCGGCGACCGTTCGAGGACTGCACTGCGTACATCTTCTGCACGACTTCGT 201
QY 124 ACGGTACATCGAATATACACGAGATGTCCTGTTGAATAGATCACTAATGCCTTAA 183
DB 202 ACGGTACATCGAACAACACACGAGATGTCCTGTTGAATGGAGCCACTGGTGCCTTGA 261
QY 184 GCATCGGTACTCCGTAGGTACATCTTCTTCTTATTTGTGCATATTTTATTGTTG 243
DB 262 GCATCGGTCCCTCCGCTGGGTACATCTGTTCTTCTGTTATTTGTGCAT-GTTTCATGCTG 320
QY 244 TTTACTGATTATACGAGTAGTTATACATACATACATACATACATACATACATACATACACA 303
DB 321 TTTACTGCTTATGCGAGTAGTTATACACATATGCACATACATGTCATCACATATATACGA 380
QY 304 ATATTTTCTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 351
DB 381 CTGATTTATCTGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 428

RESULT 3
AAF81468/G
ID AAF81468 standard; DNA; 1888 BP.
XX
XX AAF81468;
XX
```

DT 08-JUN-2001 (first entry)  
XX Corn promoter clone #700342364.  
DE Corn; promoter; transgenic plant; herbicide resistance; ds.  
KW Zea mays.  
XX WO200119976-A2.  
XX 22-MAR-2001.  
XX 13-SEP-2000; 2000WO-US25078.  
XX 16-SEP-1999; 99US-0154182.  
XX (MONS ) MONSANTO CO.  
XX Anderson HM, Chay CA, Chen G, Conner TW;  
XX WPI; 2001-244796/25.  
XX Novel promoter nucleic acid sequences useful for regulating  
PT heterologous gene expression in plants, comprising regulatory sequences  
PT located upstream to plant DNA structural coding sequences  
PS Claim 1; Page 95; 101pp; English.  
XX The present invention relates to novel corn promoter sequences (see  
CC AAF81456-AAF81478). The promoter sequences are useful for conferring  
CC expression of a second polynucleotide molecule in a transgenic plant  
CC tissue. In addition, the promoter sequences are useful for providing  
CC plants with herbicide resistance. The promoter sequences are suitable for  
CC selectively modulating expression of any operatively linked gene and  
CC provide additional regulatory element diversity in a plant expression  
CC vector in gene stacking approaches. The present sequence is one such corn  
CC promoter sequence isolated in the present invention.  
XX  
SQ -Sequence 1888 BP; 569 A; 443 C; 423 G; 453 T; 0 other;  
Query Match 9.5%; Score 124.6; DB 22; Length 1888;  
Best Local Similarity 63.9%; Pred. No. 4.6e-22;  
Matches 227; Conservative 0; Mismatches 114; Indels 14; Gaps 2;  
QY 78 CCGTTCGAGTGACTGCGGCGTACATCTTTCTGTCATCGACTTTGTACGGCTACATCGAA 137  
DB 784 CTGTTTGAGGAGTGCACCTGCTAGACCTGCTGTCATCGAGTTCGTACATTTATATCGAA 725  
QY 138 CATATACACGAGATGCTCTGTCGAATAGAGTCACTAATGCCCTTAAGCATCGGTTACTCC 197  
DB 724 CATACACACGAGATGCTCTGTCGAATGCGGCGCTGATACCTTTGAGCATCGGTCCTTC 665  
QY 198 GTAGGGTACATCTGTTCTTTCTTATTGTGTCATATTTTATTGTTTACTGATATATAC 257  
DB 664 ACAGGGAACCTCCTGTTTCCATCTTT--TGGGTGCTCTTTTGTGACTGCTTATAC 607  
QY 258 GAGTAGTATATACATACATGACATATATATATATATATATATATATATATATATATAT 317  
DB 606 GATAGTGGT-----ATACATGCTGTCACATATATATATATATATATATATATATAT 559  
QY 318 TAAATTAATAACTAAATGACTAAATTTCTAACCACCAACGACATTTGTTTCTTCCA 377  
DB 558 TAAATTAATAACTAGTATGCTCATTTCTAACATATTTTCACAGTAATAAAGGTTAGGTA 499  
QY 378 ACACTTTACCTATCTACATGTTCTTATTTCAATTTTCACTCTATAAACAACAT 432  
DB 498 TCAGCTAATTTCTTTGTTGATAGGTTATAATGACGATGCTCTATATTTATCATGAT 444  
RESULT 4  
AAS45444 standard; DNA; 6327 BP.  
XX

AC AAS45444;  
XX 18-DEC-2001 (first entry)  
XX Chemically pretreated genomic DNA associated with cell cycle #75.  
DE Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
XX PCR primer.  
XX Homo sapiens.  
OS WO200168911-A2.  
XX 20-SEP-2001.  
XX 15-MAR-2001; 2001WO-EP02945.  
XX 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX (EPIC-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2001-602751/68.  
XX Designing primers and probes for analysing diseases associated with  
PT cytosine methylation state e.g. arthritis, cancer, aging,  
PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle  
XX Claim 1; SEQ ID No 149; 28pp; English.  
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CpG dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
SQ Sequence 6327 BP; 1833 A; 93 C; 1407 G; 2994 T; 0 other;  
Query Match 3.8%; Score 50.2; DB 22; Length 6327;  
Best Local Similarity 60.7%; Pred. No. 0.013;  
Matches 82; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 211 TGTCTCTTCTTATTTGTCATATTTTATTTGTTTACTGATTATACGAGTAGTTATACA 270  
DB 2657 TATTCGTGATAGTTGGTTGAATTTTGTAGTTTGTGTTTATATATATATATATA 2716  
QY 271 TACATGCACATACATATCATATATATATATATATATATATATATATATATATA 330  
DB 2717 TAT 330  
QY 331 AAAATGACTAAATTT 345  
DB 2777 TATATAATAAAGTT 2791

RESULT 5  
 ID ABL32458/c  
 DB ABL32458 standard; DNA; 10133 BP.  
 XX  
 AC ABL32458;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 431.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytotstatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 PS  
 PS Claim 1; SEQ ID NO 431; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 XX Sequence 10133 BP; 3075 A; 70 C; 2051 G; 4937 T; 0 other;  
 Query Match 3.8%; Score 50; DB 24; Length 10133;  
 Best Local Similarity 51.0%; Pred. No. 0.018;  
 Matches 149; Conservative 0; Mismatches 135; Indels 8; Gaps 1;  
 QY 219 TTATTTGTCATATTTTATTTGTTTACTGATTATACGAGTAGTTATACATACATGCA 278  
 DB 1416 TTAATCAAAATCTCTAACTAACTCAACTCAACTAACTAACTCACTATATATACA 1357  
 QY 279 CATACATATCATCATATATACATATTTTCTTAAATTAATTAATTAATTAATGAC 338  
 DB 1356 TATCAAAATCTCAATATTTCAATATACATATATAATTAATTAATTAATTAAT 1297  
 QY 339 TAAATTTCTACACACGACATGTAATGTTTCTTCCCAACACTTTACCTATTC----- 393  
 DB 1296 TAAATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1237  
 QY 394 --TACATTTGTTCTATTTTCAATTTCTACTCTATAAACAACATAGTCTACATATGGAACA 450  
 DB 1236 TCTTAAATTTATATATCTTACAAATCTATAAATAAACAATATACATACATAAATAA 1177

QY 451 GTGCTTTGTACGACTATATACGCGATGTGTGGCTACACATAAGACAATATA 502  
 DB 1176 CTTCAATTTAATATATATATCAAAATTCCTACCATATACCTAATAATAAAAA 1125  
 RESULT 6  
 ID ABL34124/c  
 DB ABL34124 standard; DNA; 73334 BP.  
 XX  
 AC ABL34124;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 2097.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytotstatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 PS  
 PS Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 XX Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;  
 Query Match 3.8%; Score 49.4; DB 24; Length 73334;  
 Best Local Similarity 53.3%; Pred. No. 0.06;  
 Matches 104; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
 QY 252 TTATAGGAGTAGTTATACATATACATATATATATATATATATATATATATATAT 311  
 DB 59235 TTAT 59176  
 QY 312 CTAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 371  
 DB 59175 ATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 59116  
 QY 372 TCTCCAACAACCTTTACCTATTTCTACATTTCTTTTCTTCTTCTTCTTCTTCTTCTTCT 431

```
Db 59115 CAATTAATAAATAAATAATATACACACACATTCCTTAACATATCAAAAAAATCATCA 59056
QY 432 TAGCTACATGCAA 446
Db 59055 TCATCTATCATATAA 59041

RESULT 7
ID ABL34241/C
XX ABL34241 standard; DNA; 11691 BP.
AC ABL34241;
XX
DT 26-MAR-2002 (first entry)
DE Human immune system associated gene SEQ ID NO: 2214.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

Query Match 3.8%; Score 49.2; DB 24; Length 11691;
Best Local Similarity 51.1%; Pred. No. 0.03;
Matches 140; Conservative 0; Mismatches 133; Indels 1; Gaps 1;

QY 160 TGAATAGATCACTAATGCTTAAGCATCGGTTACTCCGATAGGATCATCTGTTCTTCT 219
Db 6938 TTAATAATAACCTATTACTTTAAATATTAATAATATATATACATTAATATATACCA 6879
QY 220 TATTTGTCATATTTTATTGTTGTTACTGATTATACAGTAGTATATACATACATGAC 279
Db 6878 TAATTCATAACTACATATACATTTTAATAATAACT-ATTAAATATACACAAAAAAT 6820
QY 280 ATACATATCATCATATATATCAATATTTTCTAAATTAATAATTAACCTAAATATGACT 339
```

```
Db 6819 ATATATATATACATACATATAAATAAATAAATTTTACATATATAAATACACATATAATTTT 6760
QY 340 AAATTTCTAACACCAACGACATTTGTAATGTTTCTTCCAAACACTTTTACCTATTCTACATT 399
Db 6759 ACTAATCTATAAATAAATACATATTAACACTTTTATTTTAAATATATATATAAATCTTATA 6700
QY 400 GTTCTATTTTCAATTTTCACTCTATAAACAACATA 433
Db 6699 TTTCAAAAATAATTTTACACAATAAAAAAATACA 6666

RESULT 8
AAZ17263
ID AAZ17263 standard; CDNA; 1337 BP.
XX
AC AAZ17263;
XX
DT 12-OCT-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
XX Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
XX Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types

Claim 1; Page 2250-2251; 2479pp; English:

The present invention describes a library of human polynucleotides
comprising the sequences given in AAZ12532 to AAZ1779. Also described is
a method of detecting differentially expressed genes correlated with the
cancerous state of a mammalian cell, comprising detecting at least one
differentially expressed gene product in a test sample from a cell
suspected of being cancerous, where the gene product is encoded by one
of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
polynucleotides can be used as a source of primers and probes, which can
be used for a variety of purpose, e.g. detection of expression levels,
mapping, tissue typing or profiling, forensics, genetic analysis and
detection of polymorphisms. Polypeptides encoded by the polynucleotides
can be used for raising antibodies for experimental, diagnostic and
therapeutic purposes. The polynucleotides may also be used to construct
arrays for diagnostics (which may be used to determine function of an
encoded protein); and to detect differences in expression levels between
two cells (e.g. to identify abnormal or diseased tissue in a human, to
identify a genetic predisposition or susceptibility to a disease such as
```

CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX  
SQ Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;  
Query Match 3.7%; Score 48.8; DB 20; Length 1337;  
Best Local Similarity 32.4%; Pred. No. 0.015;  
Matches 101; Conservative 0; Mismatches 211; Indels 0; Gaps 0;  
QY 755 CACTACCCCAACTCTCACCGGAAACCGCGTGGATCGGCAATCAACGAGGTGGTGC 814  
DB 350 CCNCNCNN 409  
QY 815 CCGGTGGCCACTCTCAGCTGCACGGCACCATCCCTCTGCAGCCGCTCACCAGCATGCC 874  
DB 410 CCCCCNCNCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 469  
QY 875 GTGTGCGGAAGGACACACACCCGCCCAACCCACCTCAGGAAACCCGTCGCGGCGTGCC 934  
DB 470 NNNNNNNNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 529  
QY 935 CGTGTGGTCCGCGCTCGCGCAACGAGGGGCGCGGCTGCTGAGTCCCGTGGACACCCGA 994  
DB 530 CNCCGCCGCCNCNCNCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 589  
QY 995 CACCCTGTGGCGCTTGTATTATCCCGGAAATCTCATCTGCCCCCAGCGCCGACTGC 1054  
DB 590 NCNCNCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCCNNCC 649  
QY 1055 GTGCGCGCGGCC 1066  
DB 650 CCCCCCCCCCCC 661

RESULT 9  
AAH93294/C  
ID AAH93294 standard; DNA; 399 BP.  
XX  
AC AAH93294;  
XX  
DT 04-OCT-2001 (first entry)  
XX  
DE Plasmodium falciparum MAL3P8 polynucleotide SEQ ID NO 16.  
XX  
KW Human; antisense-therapy; gene-therapy; diagnostic; forensic;  
KW gene mapping; ds.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO200152616-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 22-DEC-2000; 2000WO-US35190.  
XX  
PR 23-DEC-1999; 99US-0471275.  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-451890/48.  
XX  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
PT antibodies and research use -  
XX  
XX Example 4; Page 106; 135pp; English.  
PS  
XX The invention relates to an isolated human polynucleotide (AAH75398)  
CC

CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and  
CC gene-therapy, in diagnostics, forensics, gene mapping and identification  
CC of mutations responsible for genetic disorders and other traits.  
CC Polynucleotide sequences with potential homology were also identified  
CC (AAH93283-AAH93356).

XX  
SQ Sequence 399 BP; 225 A; 24 C; 15 G; 135 T; 0 other;  
Query Match 3.7%; Score 48; DB 22; Length 399;  
Best Local Similarity 47.1%; Pred. No. 0.014;  
Matches 147; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 208 TTCGTCTCTCTTATTTGTCATATTTTATTTCTGTTTACTGATTATACGAGTACTTAT 267  
DB 320 TTTTITTTTTTATATATTTGTCATATTTTATTTGTCATTAACGATATATATATCTTT 261  
QY 268 ACATACATGCACATACATATCATCATATATATATATATATATATATATATATATATAT 327  
DB 260 TTTTACAT 201  
QY 328 CTAAAAATGACTAAATTTCTAACACCAACACATTTGTAATGTTTCTCCCAACACTTAC 387  
DB 200 ATATATATTTTATTTTATTTTACGATAGTTTAAATTTTTTTTTTTTATCTCGCAATTA 141  
QY 388 CTATCTACATTTGTTCTATTTTCGAATTTCACTCTATAAACAACATAGTCTACAAATGAAA 447  
DB 140 AACATTTTAAATTTGTTCTTTATTTATTAATTAATATATATATATATATATATATATAT 81  
QY 448 ACAGTGTCTTTGACGACTATATACGCGATGTGGGCTACACACATAAGACAAATAGTCGT 507  
DB 80 TTAGAATGTTTGTATATTTTAAAGGTTTTTTTTTTTTTTTTTTTTTTTGATACACAATTCATTA 21  
QY 508 TTGAAGATTGAA 519  
DB 20 TTTAAATTATAA 9

RESULT 10  
AAS56505/C  
ID AAS56505 standard; cDNA; 326 BP.  
XX  
AC AAS56505;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human cDNA for an ovarian cancer protein #129.  
XX  
KW Human; ss; ovarian cancer protein; cancer; tumour; ovarian cancer;  
KW endometrial cancer; cytostatic.  
XX  
OS Homo sapiens.  
XX  
PN WO200170976-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 20-MAR-2001; 2001WO-US09062.  
XX  
PR 21-MAR-2000; 2000US-190710P.  
PR 22-JUN-2000; 2000US-213748P.  
PR 19-DEC-2000; 2000US-257276P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Xu J, Pyle RA, Stolk JA;  
XX  
DR WPI; 2001-607531/69.  
XX  
XX Nucleic acids encoding 222 polypeptides associated with ovarian and  
PT endometrial cancers, useful for diagnosing, preventing and treating  
PT cancers -  
XX  
XX Claim 1; Page 157; 187pp; English.  
PS

The invention relates to human polynucleotides encoding proteins associated with ovarian and endometrial cancers. The polynucleotides and the proteins they encode may be used in the prevention, diagnosis and treatment of diseases associated with the inappropriate expression of ovarian and endometrial cancer polypeptides (OECRs). For example, the polynucleotide (or an expression vector comprising the polynucleotide) and the OECR may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of OECRs by expressing inactive proteins or to supplement the patients own production of them. Additionally, the polynucleotide may be used to produce the OECRs, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The polynucleotide and its complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The OECRs may also be used as antigens in the production of anti-OECR antibodies and in assays to identify modulators of it's expression and activity. The anti-OECR antibodies and antagonists may also be used to down regulate expression and activity. The anti-OECR antibodies may also be used as diagnostic agents for detecting the presence of OECR in samples (e.g. by enzyme linked immunosorbant assay (ELISA)) and hence diagnose patients with cancers. The present sequence is a ovarian and endometrial cancer linked cDNA of the invention.

SQ Sequence 326 BP; 130 A; 48 C; 34 G; 114 T; 0 other;  
 Query Match 3.6%; Score 47.8; DB 22; Length 326;  
 Best Local Similarity 55.0%; pred. No. 0.014;  
 Matches 94; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY	211	TGTTCTTCTTATTTGTCATATATTTTATTGTTGTTTACTGATATATACGAGTAGTTATACA	270
Db	326	TTTTTTTTTTTTTTTTTTTTTTTTCAGGGATTGAACACATTTTAAATTATTAGAAATACA	267
QY	271	TACATGCACATACATATCATCATATATATCACAATATTTTCTCAAATTTAAATTAACAACTA	330
Db	266	TACATGCATATAATTAACCTTAATTTATTTTAAAAATAAATTTATGGTCATGAAGAAACA	207
QY	331	AAATGACATAAATTTCTACACCAACGACATGTTAATGTTTCTCCACAA	381
Db	206	ATAGTGCTTAAATTCGAATTTTCATAGACATGTAATTTTATTCCTTAAAAA	156

RESULT 11  
ABL34125/c  
ID ABL34125 standard; DNA; 73334 BP.

XX	ABL34125;
AC	
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 2098

Human; immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antianaemic; cytostatic; nootropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmologic; antirheumatic; antiarthritic; antidiabetic; antipsoriatic; antinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

XX	Homo sapiens.
OS	WO200200928-A2.
XX	
XX	03-JAN-2002.
PN	02-JUL-2001; 2001WO-EP07537.
XX	
XX	30-JUN-2000; 2000DE-1032529.
XX	
PR	

01-SEP-2000; 2000DE-1043826.  
(EPIG-) EPIGENOMICS AG.  
Olek A. Piepenbrock C, Berlin K;  
WPI: 2002-130909/17.  
Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation -  
Claim 1; SEQ ID NO 2098; 32pp + Sequence Listing; German.  
The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.  
Sequence 73334 BP; 24605 A; 858 C; 14432 G; 33439 T; 0 other:

QY	292	ACATATATCACAAATATTTTTCTTAAATTAAATTTAAACTAAATAAATGACTAAATTTCTTAACA	351
Db	66218	AAAAATTTACATATTTTTTAATAAATAATACAAAATTAATAAAAAAACAATTTCAATA	66159
QY	352	CCAAAGCAGATGTGAATGTTTTCTCCACAACACTTTACCTATTCTACATGTTCTTATTTTCTGA	411
Db	66158	ACAACAAACATAATACTTTATACCTATTATCCCAACACTTTTAAAAAAGCAAAACGAAA	66099
QY	412	ATTTCACCTCTATAACAACATAGTCTACAATGGAACA	450
Db	66098	AAATCACTTAATCCCAAAAATTCGAATCCCAACGTAACA	66060

RESULT 12  
ABL33574/C  
ID ABL33574 standard; DNA; 6627 BP.

XX	ABL33574;
XX	AC
XX	DT
XX	26-MAR-2002 (first entry)
DE	Human immune system associated gene SEQ ID NO: 1547

Human; immune system disease; cytosine methylation; antiasthmatic;  
antiarteriosclerotic; anti anaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritis; antidiabetic; antiparasitic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene: ds.

XX Homo sapiens.  
OS  
XX  
XX  
PN WO200200928-A2.  
XX  
XX  
XX  
PD 03-JAN-2002.  
XX  
XX  
PF 02-JUL-2001; 2001WO-EF07537.  
XX

```
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX Claim 1; SEQ ID NO 1547; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 6627 BP; 2052 A; 117 C; 1386 G; 3072 T; 0 other;
SQ
Query Match 3.6%; Score 47.4; DB 24; Length 6627;
Best Local Similarity 49.2%; Pred. No. 0.069; 4; Gaps 1;
Matches 155; Conservative 0; Mismatches 156; Indels 4;
QY - 192 TACTCGTAGGTFACATTCCTGTTCTTCTTATTGTCATATTTTATTGTTTACTCA 251
DB 6510 TACTCTTTAAATACCTACTAATCTCAATTCCTTCACTTAACCTAAATAA 6451
QY 252 TTATACAGTAGTTATACATATGCGATACATATCATCATATATACATATTTT 311
DB 6450 TTCAATATATTAATTTTCAATAAATAAACTAAACCAAAATTAATAACTTTA 6391
QY 312 CTAAAT---AAATTAACCTAAATGACATAATTTTCAACACACGACATGTAAT 367
DB 6390 CTAACTTTACCCCACTAAATAAATAAACTAAATTTTAAACCCCAAACTAAACCAT 6331
QY 368 GTTTTCTCGAACACTTTACCTATTCATATGTCATATTTTGAATTTTCACTCTATAAC 427
DB 6330 ATAATCTAAACCACTTAACTCCATAATCACTAAACATAAATCAATTAATCTCT 6271
QY 428 AACATAGTCTACAAATGAAACAGTCTTTGTGACGACTATATACGGGATGTGCGCTACA 487
DB 6270 ATACAACTCCAAACGCAACAACTACTATCTCCATTTTCAAAACCAACAACTAAACA 6211
QY 488 ACATAAGACAATATA 502
DB 6210 CCAATCGAAACCTA 6196
RESULT 13
ID AAA70152/c
XX AAA70152 standard; DNA; 6033 BP.
XX
XX AAA70152;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
XX Plasmodium falciparum.
OS
XX WO200025728-A2.
XX
XX 11-MAY-2000.
```

```
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 493-495; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins),
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 other;
SQ
Query Match 3.6%; Score 47.2; DB 21; Length 6033;
Best Local Similarity 49.6%; Pred. No. 0.074;
Matches 121; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 205 ACATTCCTCTCTCTTCTTATTTGTCATATTTTATTGTTGTTTACTGATTATACGAGTAGT 264
DB 2523 ACAATCTTTTCTATGTACACCATATTTATATAATTTATATATATATATATATATAT 2464
QY 265 TATACATACATGCACATACATATCATCATATATATACAAATATTTTCTTAATAAATTA 324
DB 2463 GTTAAATGTAATATCAATATTTTCAACATTTTATTTAAAAATTTTAAAAA 2404
QY 325 AAACATAAAATGACTAAATTTCTAACACACGACATTTGTAATGTTTCTCCACAACCTT 384
DB 2403 TTTTGAAAAATCAATCAATATATATTTTCATATATAAATTTGAAAGATTTATACATCCA 2344
QY 385 TACCTATTCTACATGTTCTTATTTTGAATTTTCACTCTCTATAAACAACATAGTCTACAATGG 444
DB 2343 TTCTCTTAATCCTTTTTCATTTTGTGCTTTTATTTTATATATATATATATATATG 2284
QY 445 AAAA 448
DB 2283 GAAA 2280
RESULT 14
ID AAL16482
XX AAL16482 standard; cDNA; 281 BP.
XX
```







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:44:03 ; Search time 39.0149 Seconds  
(without alignments)  
8247.617 Million cell updates/sec

Title: US-09-905-558C-3  
Perfect score: 1310  
Sequence: 1 cccatcgctgtttgtctac.....caagggaagatgatcgatg 1310

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	3.9	7218	1	US-08-232-463-14
2	45.2	3.5	4257	2	US-08-690-473-1
3	45.2	3.5	4257	4	US-09-259-821A-1
4	45.2	3.5	4257	4	US-08-843-659-1
5	45.2	3.5	12001	1	US-08-458-568A-11
6	43.4	3.3	2781	3	US-08-749-522-4
7	43	3.3	796	4	US-09-007-119-14
8	43	3.3	809	4	US-09-007-119-10
9	41.8	3.2	1890	6	5312912-3
10	41.6	3.2	1441	4	US-08-821-994-63
11	41	3.1	4254	2	US-08-443-639-7
12	40.6	3.1	6152	4	US-08-973-462-1
13	40.4	3.1	1000	1	US-08-599-252-96
14	40.4	3.1	1000	5	PCT-US96-06352-96
15	40.4	3.1	1000	5	PCT-US96-06583-96
16	40.4	3.1	10342	4	US-08-372-927-5
17	40	3.1	4673	1	US-07-638-431-1
18	40	3.1	4673	5	PCT-US92-00018-1
19	40	3.1	5829	4	US-09-004-838-109
20	39.6	3.0	5192	3	US-08-947-823-1
21	39.2	3.0	19124	2	US-08-487-826B-13
22	39	3.0	320	4	US-09-165-264-11
23	38.8	3.0	923	4	US-09-004-731-66
24	38.8	3.0	923	4	US-08-749-699-66
25	38.8	3.0	152331	3	US-09-128-155-16
26	38.6	2.9	320	4	US-09-165-264-13
27	38.4	2.9	1368	3	US-08-874-563-5

c 28	38.4	2.9	1368	3	US-08-577-483-14	Sequence 14, Appl
c 29	38.4	2.9	1736	3	US-09-182-816-22	Sequence 22, Appl
c 30	38.4	2.9	1736	3	US-09-182-816-24	Sequence 24, Appl
c 31	38.4	2.9	1736	3	US-09-471-528-22	Sequence 22, Appl
c 32	38.4	2.9	1736	3	US-09-471-528-24	Sequence 24, Appl
c 33	38.4	2.9	1736	4	US-09-634-530-22	Sequence 22, Appl
c 34	38.4	2.9	1736	4	US-09-634-530-24	Sequence 24, Appl
c 35	38.4	2.9	4253	3	US-08-577-483-7	Sequence 7, Appl
c 36	38.2	2.9	4526	1	US-07-855-412B-4	Sequence 4, Appl
c 37	38.2	2.9	4526	2	US-08-308-887A-4	Sequence 4, Appl
c 38	38.2	2.9	4526	3	US-08-881-094-4	Sequence 4, Appl
c 39	38	2.9	319	4	US-09-165-264-8	Sequence 8, Appl
c 40	37.8	2.9	611	4	US-09-385-982-393	Sequence 393, App
c 41	37.8	2.9	899	1	US-07-820-154A-3	Sequence 3, Appl
c 42	37.8	2.9	899	2	US-08-097-554A-3	Sequence 3, Appl
c 43	37.8	2.9	899	3	US-08-480-640A-3	Sequence 3, Appl
c 44	37.8	2.9	899	3	US-08-295-802-3	Sequence 3, Appl
c 45	37.8	2.9	899	4	US-08-488-237A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
: Sequence 14, Application US/08232463  
: Patent No. 5670367  
: GENERAL INFORMATION:  
: APPLICANT: DORNER, F.  
: APPLICANT: SCHEIFLINGER, F.  
: APPLICANT: FALKNER, F. G.  
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
: NUMBER OF SEQUENCES: 52  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Foley & Lardner  
: STREET: 1800 Diagonal Road, Suite 500  
: CITY: Alexandria  
: STATE: VA  
: COUNTRY: USA  
: ZIP: 22313-0299  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/232,463  
: FILING DATE:  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US/07/935,313  
: FILING DATE:  
: APPLICATION NUMBER: EP 91 114 300.6  
: FILING DATE: 26-AUG-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (703)836-9300  
: TELEFAX: (703)683-4109  
: TELEX: 899149  
: INFORMATION FOR SEQ ID NO: 14:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 7218 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: IMMEDIATE SOURCE:  
: CLONE: ptr9pt-Fls  
US-08-232-463-14

Query Match 3.9%; Score 51.2; DB 1; Length 7218;

[illegible]

```

; Sequence 1, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopardi, Rosario
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND TCP4 AS
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,659
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARSB:519
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4257 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-843-659-1

```

Query Match	3.5%;	Score 45.2;	DB 4;	Length 4257;
Best Local Similarity	49.3%;	Pred. No. 0.0063;		
Matches 145;	Conservative 0;	Mismatches 148;	Indels 1;	Gaps 1;
QY 714	CAGGCTCGCTGGCCCCACGGCGTGCTGTCAGGAAGGCGACTACCCCAACCTTCAC	773		
Db				
2513	CCGCCTTCGCGCCGCGGGAGGGGGCAAGCAAGAGTCCGGGCCCGCGCGCGC	2572		
QY 774	CGAAAAACCGGGCTGGATCGGCAATCAACAGAGGTGGTGCCTCCGTCGCCCACTCTCCACG	833		
Db				
2573	CCGAGGGGGCGGGCGCGGACCCCGAAGAGAGAGAGGGCGGACGCCGCCCGGCT	2632		
QY 834	TCCACGGCACCATCCCTCTGCGAGCGGCTCACACGCCATCGCTGCGGGGAACGGCACAA	893		
Db				
2633	CGGAGCCCGCGGCCCTCCCGCGCGCGCGGCCCTCGACGCCCGCGGGGCCCGAGC	2692		
QY 894	CCACCCCCAACCCACTCAGCAAAACCCGTCGCCGCCGTGCGCGCTGCGGTCGCGGCTCGG	953		
Db				
2693	CCGCCCCCGCCAGCCG	2752		
QY 954	CAA-CGAGGGGGCCCGCGCTGCTAGTCCCTGGACACCCGACACCCCTGTGCGC	1006		
Db				
2753	TGGCCGTGTGCGCGCGGCCGCCGAGGGCCCCGACCCCTGGGCGGCTGGCGCG	2806		

```

RESULT 5
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
; GENERAL INFORMATION:
; APPLICANT: Schaffer, Priscilla A.
; APPLICANT: Yeh, Lily
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus
; TITLE OF INVENTION: Infections

```

```

1  NUMBER OF SEQUENCE: 15
2  CORRESPONDENCE ADDRESS:
3  ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339rls
4  STREET: One Liberty Place, 46th floor
5  CITY: Philadelphia
6  STATE: PA
7  COUNTRY: USA
8  ZIP: 19103
9  COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: WordPerfect 5.1
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/458,568A
16 FILING DATE: 02-JUNE-1995
17 CLASSIFICATION: 435
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 08/065,146
20 FILING DATE: 05-MAY-1993
21 CLASSIFICATION: 435
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Leary Ph.D., Kathryn R.
24 REGISTRATION NUMBER: 36,317
25 REFERENCE/DOCKET NUMBER: DFCI-0029
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (215) 568-3100
28 TELEFAX: (215) 568-3439
29 INFORMATION FOR SEQ ID NO: 11:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 12001 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 HYPOTHETICAL: NO
37 ANTI-SENSE: NO
38 ORIGINAL SOURCE:
39 ORGANISM: Herpes simplex virus
40 STRAIN: Herpes simplex virus Type 1
41 US-08-458-568A-11

```

```

Query Match      3.5%; Score 45.2; DB 1; Length 12001;
Best Local Similarity 49.3%; Pred. No. 0.011;
Matches 145; Conservative 0; Mismatches 148; Indels 1; Gaps 1;

Qy 714 CAGGCTCGCTGGCCCCACGGCGTGCTGCTGACGAAGGGCACTACCCCAACCTCTCAC 773
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4161 CCGCTTCGGCCGCGCGGGAGGGGCGAAGCGAGAGTCCCGGCCCGCCCGCGCGC 4102

Qy 774 CGAATAACCGCGCTGGATCGGCAATCAAAACGAGGTGGTCCCGCTGCCCACTCTCCACG 833
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4101 CCGGAGGGCGGGCGCGCCGACCCCGAAGACGAAGAAGAGGGCGCGACGCCCCCGCT 4042

Qy 834 TCCACGCGACATCCCTCTGCGAGCGCTCACCAGCCATGCGTGTGCGGGNACGGCACAA 893
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4041 CGGACGCGCGGGCGCCCTCTCCCGCGCGCGCCCCCTCCACGCCCGCGGGGCCGAGC 3982

Qy 894 CCACCCCCAACCCCACTCACGAAACCCCGCTGCCGGCCGTGCCCGTGTGCGGTCCGGCTCGG 953
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3981 CCGCCCCCGCCACGCCCGCGCGCGCCCCCGGGCGCGCGGCGCAGGCCCGCCCGCGCCCCG 3922

Qy 954 CAA-CGAGGGGGCGCGCTGTGAGTCCCTGGACACCCGACACCCCTGTGCGC 1006
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3921 TGGCCGTGTGCGCGCCCGCCCGCGAGGGCGCCGACCCCTGGGGCGGTGGCGCGC 3868

RESULT 6
US-08-749-522-4/c
; Sequence 4, Application US/08749522
; Patent No. 6096950
; GENERAL INFORMATION:
; APPLICANT: John, Mallivakal

```





Tue Nov 12 13:19:27 2002

```

Query Match          3.2%; Score 41.6; DB 4; Length 1441;
Best Local Similarity 48.7%; Pred. No. 0.037;
Matches 113; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 211 TGTCTCTCTTATTTGTGCATATTTTTTATTTGTTGTTTACTGATTATACGAGTAGCTGTTATACA 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1441 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1382

QY 271 TACATGCCACATACATCATCACATATATACATATATTTTCTTAAATTTAAATTTAAACATA 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1381 TTTT TTTGAAAATATATCATTTTATTTATATATCTCTTAATACATATCAAGCCCAAGCA 1322

QY 331 AAAATGACTAAATTTCTTAACACCAACGACATGTAAATGTTTTCCTCAACAACTTTTACCCTA 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1321 AATATTTCTTTGAATGAAATGATTTTGACCAACATAGAGATTTATACAGATTTAAATCCCTA 1262

QY 391 TTCTCAATGTTCTATTTTCGAATTTTCACATCTCTATAACAACTATCTACAT 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1261 TGGTGAATCAATTTTTTTCGAAATACTATACATACAAATAATAACAAGT 1210

RESULT 11
US-08-443-639-7/c
; Sequence 7, Application US/08443639
; Patent No. 5981843
; GENERAL INFORMATION:
; APPLICANT: Chappell, Joseph
; APPLICANT: Yin, Shaohui
; APPLICANT: Cornett, Catherine A.G.
; TITLE OF INVENTION: Transcriptional Control Sequences and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/443,639
; FILING DATE: 18-MAY-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 69-94
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4254 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(1217..1327, 1455..1718, 1806..2182,
; LOCATION: 2259
; LOCATION: .12477, 2609..2747, 2903..3148, 3262..3558)
US-08-443-639-7

```

```

Query Match          3.1%; Score 41; DB 2; Length 4254;
Best Local Similarity 52.7%; Pred. No. 0.098;
Matches      89; Conservative    0; Mismatches   80; Indels     0; Gaps     0;

QY  265 TATACATACATGCACATACATATCATCACAATATATCACAAATATTTTCTTAATAATTAAATTA 324
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  367 TGTACAGACACACACACACATACATTATATATATATATATATATATATATATATATATATA 308
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  325 AAACCTAAAAAAGTACTAAATTTCTAACACCAGACGATTTGAATGTTTTCTCCCAACAACCTT 384
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  307 TATATATATATATATATATATATTAATTTCTATAATTTTCTTAACATTTTCACAACTTTT 248
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  385 FACCTATTCTCAATGGTTCTATTTCGAAATTTCTACTCTATAAACAACATA 433
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  247 AGTGATTTTTCTCGATTATATAAATGATTTTATAAATTTATAAANATATA 199
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRAULHE, PIERRE
; APPLICANT: DAUBERSTES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatenLin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

Query Match          3.1%; Score 40.6; DB 4; Length 6152;
Best Local Similarity 54.3%; Pred. No. 0.16;
Matches      82; Conservative    0; Mismatches   69; Indels     0; Gaps     0;

QY  199 TAGGTGACATTCGTCTCTTTCTTTATTTGTGCATATTTTATTTGTTTGTGTTTACTGATATACG 258
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  6009 TAAGCTAAATTTTATCTTCTTAATATATATACATATATATATATATATATATATATATATAAT 5950
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  259 AGTAGTTATACATACATGCACATACATATCATCATATATATACAAATATTTTCTCAAATTT 318
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  5949 ATAAATTTTAATAAAAACAAATAATTAATATATACATATATCGTGAATAATATACAGT 5890
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  319 AAATTTAAACTAAAAATGACTAAATTTCTAA 349
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  5889 TTCTTACAGAAATCTATATATTAATAAATAA 5859
Db   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-08-599-252-96
; Sequence 96, Application US/08599252
; Patent No. 5705343
; GENERAL INFORMATION:
; APPLICANT: DRAYNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIRKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER

```



STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06352  
FILING DATE: 09-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-599-252-96

Query Match 3.1%; Score 40.4; DB 1; Length 1000;  
Best Local Similarity 46.4%; Pred. No. 0.067;  
Matches 122; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY - 208 TTCTGTTCTTCTTATTTGTCATATTTTATTTGTTTACTGATATATAGGAGTAGTTAT 267  
DB 117 TTATTTATTTTATTTAAATAATATTTAAATATTTATTTATTTACTTATATAT 176  
QY 268 ACATACATGCACATACATATATATATATATATATATATATATATATATATATAT 327  
DB 177 TTATTTATTTTATTTAAATAATATTTATATATATATATATATATATATATATAT 236  
QY 328 CTAATAATGACTAAATTTCTAACACCAAGCATTTGTAATGTTTCTTCCACAACTTTAC 387  
DB 237 TGTCACTCTATTTTAAATATATATATATATATATATATATATATATATATATAT 296  
QY 388 CTATTTCTACATTTGTTCTATTTTGAATTTTCACTCTATAAAACACATAGTCTACAATGGA 447  
DB 297 GTTTTCCTCTACTGTGTAATAAGACATACTCACTAAGGGGAAGATCTTGGCCCNAG 356  
QY 448 ACAGTGCTTTGTACGACTATATA 470  
DB 357 TGTGTGATAAATCANANANANA 379

RESULT 14  
PCT-US96-06352-96  
Sequence 96, Application PC/TUS9606352  
GENERAL INFORMATION:  
APPLICANT: DRAYNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIRKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC

COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06352  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06352-96

Query Match 3.1%; Score 40.4; DB 5; Length 1000;  
Best Local Similarity 46.4%; Pred. No. 0.067;  
Matches 122; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 208 TTCTGTTCTTCTTATTTGTCATATTTTATTTGTTTACTGATATATAGGAGTAGTTAT 267  
DB 117 TTATTTATTTTATTTAAATAATATTTAAATATTTATTTATTTACTTATATAT 176  
QY 268 ACATACATGCACATACATATATATATATATATATATATATATATATATATATAT 327  
DB 177 TTATTTATTTTATTTAAATAATATTTATATATATATATATATATATATATATAT 236  
QY 328 CTAATAATGACTAAATTTCTAACACCAAGCATTTGTAATGTTTCTTCCACAACTTTAC 387  
DB 237 TGTCACTCTATTTTAAATATATATATATATATATATATATATATATATATATAT 296  
QY 388 CTATTTCTACATTTGTTCTATTTTGAATTTTCACTCTATAAAACACATAGTCTACAATGGA 447  
DB 297 GTTTTCCTCTACTGTGTAATAAGACATACTCACTAAGGGGAAGATCTTGGCCCNAG 356  
QY 448 ACAGTGCTTTGTACGACTATATA 470  
DB 357 TGTGTGATAAATCANANANANA 379

RESULT 15  
PCT-US96-06583-96  
Sequence 96, Application PC/TUS9606583  
GENERAL INFORMATION:  
APPLICANT: DRAYNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GNIRKE, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC

COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06583  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06583-96

Query Match 3.1%; Score 40.4; DB 5; Length 1000;  
Best Local Similarity 46.4%; Pred. No. 0.067;  
Matches 122; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 208 TTCTGTTCTTCTATTGTGTCATATTTTATTTGTTGTTTACTGATTATACGAGTAGTTAT 267  
Db 117 TTATTATTATTATAAAATAATTTAATAATTAATTTATTTACTTATTTAT 176  
QY 268 ACATACATGCACATACATATCATCATATATCACAATATTTTCTAAATTAATAAAA 327  
Db 177 TTATTATTATAAATTTATAATATGTCACATTATAAAATATTTAATTAATAATTTAA 236  
QY 328 CTAAAAATGACTAAATTTCTACACCAAGCATGTAATGTTTCTCCACAACCTTAC 387  
Db 237 TGTCACTCTATTATTATATCATAAATACATAAATACATTTTCTGTAGAAATCAC 296  
QY 388 CTATTTACATGTTCTATTTCGAATTTTCACTCTATAAACAACATAGTCTACAATGGAAA 447  
Db 297 GTTTCCTCTACTGTGAATAAGACATAAATCACTCACTAAGGGGAAGAATCTTGGCCCNAA 356  
QY 448 ACAGTGCCTTTGTACGACTATATA 470  
Db 357 TGTGTATAAATCANANANANA 379

Search completed: November 7, 2002, 10:47:55  
Job time : 105.015 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 08:30:58 ; Search time 1237.97 Seconds  
(without alignments)  
14282.232 Million cell updates/sec

Title: US-09-905-558c-3  
Perfect score: 1310  
Sequence: 1 cccatcgctgtttgtctac.....caagggaagtgatcgatg 1310

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pin.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	149.2	11.4	374	12	BH255115
C 2	123.8	9.5	223	12	A2921021
C 3	107.6	8.2	496	9	AW928053
C 4	58.2	4.4	572	10	BI885732
C 5	58.2	4.4	1101	12	CNS0106X
C 6	57.8	4.4	500	9	AU087781
C 7	57.4	4.4	935	12	CNS006XK
C 8	56.6	4.3	534	12	BH255739
C 9	55.6	4.2	1058	12	CNS014D0
C 10	55.2	4.2	818	10	BI952390
C 11	55	4.2	925	12	CNS0091P
C 12	54.8	4.2	936	10	EG852371
C 13	54	4.1	798	12	AQ251130
C 14	54	4.1	1101	12	CNS00FVE
C 15	53.8	4.1	481	12	CNS016IP
C 16	53.8	4.1	848	10	BI950718
C 17	53.8	4.1	1038	12	CNS015VV
C 18	53.4	4.1	802	12	CNS02R43
C 19	53.4	4.1	1101	12	CNS008WC
C 20	53.2	4.1	945	12	CNS04D0K
C 21	53.2	4.1	1082	10	BE455154
C 22	53	4.0	914	12	A674804
C 23	53	4.0	928	12	CNS00DKY
C 24	53	4.0	1101	12	CNS0039G
C 25	52.8	4.0	652	12	BH255629
C 26	52.8	4.0	928	12	CNS00DKY
C 27	52.6	4.0	1092	12	CNS020K7
C 28	52.6	4.0	1101	12	CNS017KE
C 29	52.6	4.0	1101	12	CNS017ZQ
C 30	52.4	4.0	479	10	BM163977
C 31	52.4	4.0	745	12	A2046300
C 32	52.2	4.0	914	12	CNS00HJY
C 33	52.2	4.0	1101	12	CNS00H18
C 34	52	4.0	861	10	BI953364
C 35	52	4.0	1101	12	CNS017EJ
C 36	51.8	4.0	649	12	BI9263
C 37	51.8	4.0	1043	12	CNS0145P
C 38	51.8	4.0	1101	12	CNS0023I
C 39	51.4	3.9	553	9	AI657313
C 40	51.4	3.9	563	10	BG004199
C 41	51.4	3.9	776	12	AGL38889
C 42	51.4	3.9	910	12	CNS006ON
C 43	51.4	3.9	925	10	BG441241
C 44	51.4	3.9	1101	12	CNS016LI
C 45	51.2	3.9	342	10	BM168596

C 18	53.4	4.1	802	12	CNS02R43	AL210108 Tetraodon
C 19	53.4	4.1	1101	12	CNS008WC	AL052719 Drosophila
C 20	53.2	4.1	945	12	CNS04D0K	AL285149 Tetraodon
C 21	53.2	4.1	1082	10	BE455154	BE455154 HVSMEH009
C 22	53	4.0	914	12	A674804	AZ674804 ENTKN64TR
C 23	53	4.0	928	12	CNS00DKY	AL071865 Drosophila
C 24	53	4.0	1101	12	CNS0039G	AL063921 Drosophila
C 25	52.8	4.0	652	12	BH255629	BH255629 LDH5BAM00
C 26	52.8	4.0	928	12	CNS00DKY	AL071865 Drosophila
C 27	52.6	4.0	1092	12	CNS020K7	AL175896 Tetraodon
C 28	52.6	4.0	1101	12	CNS017KE	AL108152 Drosophila
C 29	52.6	4.0	1101	12	CNS017ZQ	AL108704 Drosophila
C 30	52.4	4.0	479	10	BM163977	BM163977 EST566500
C 31	52.4	4.0	745	12	A2046300	AZ046300 nbeh0091L
C 32	52.2	4.0	914	12	CNS00HJY	AL097768 Drosophila
C 33	52.2	4.0	1101	12	CNS00H18	AL072804 Drosophila
C 34	52	4.0	861	10	BI953364	BI953364 HVSMEH001
C 35	52	4.0	1101	12	CNS017EJ	AL107941 Drosophila
C 36	51.8	4.0	649	12	BI9263	BI9263 T22C5-T7 TA
C 37	51.8	4.0	1043	12	CNS0145P	AL103735 Drosophila
C 38	51.8	4.0	1101	12	CNS0023I	AL097176 Drosophila
C 39	51.4	3.9	553	9	AI657313	AI657313 486093C05
C 40	51.4	3.9	563	10	BG004199	BG004199 QV4-GN012
C 41	51.4	3.9	776	12	AGL38889	AGL38889 Pan trogl
C 42	51.4	3.9	910	12	CNS006ON	AL065629 Drosophila
C 43	51.4	3.9	925	10	BG441241	BG441241 GA_Ea001
C 44	51.4	3.9	1101	12	CNS016LI	AL106896 Drosophila
C 45	51.2	3.9	342	10	BM168596	BM168596 EST571119

ALIGNMENTS

RESULT 1  
BH255115/c  
LOCUS BH255115 374 bp DNA linear GSS 29-NOV-2001  
DEFINITION RXL1BAM001A01r Zea mays L. XLI methyl filtration maize root  
genomic shotgun library Zea mays genomic clone RXL1BAM001A01r, DNA  
sequence.  
ACCESSION BH255115  
VERSION BH255115.1 GI:17150008  
KEYWORDS GSS.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
REFERENCE 1 (bases 1 to 374)  
AUTHORS Kim, S.W., Yu, Y., Lee, M.C., Yang, T.J., Main, D., Henry, D., Oates, R.  
and Wing, R.A.  
TITLE Genomic shotgun library from maize  
JOURNAL Unpublished (2001)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: AATTAACTCCTCACTAAAGGG  
Class: shotgun  
High quality sequence stop: 374.  
FEATURES  
source  
Location/Qualifiers  
1..374  
/organism="Zea mays"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="RXL1BAM001A01r"  
/clone\_lib="Zea mays L. XLI methyl filtration maize root  
genomic shotgun library"  
/tissue\_type="Root"  
/lab\_host="XLIBlue"  
/note="Vector: pCUGtblu-1; Site\_1: Sau3A1;



```

BASE COUNT      203 a      62 c      107 g      124 t
ORIGIN

Query Match      8.2%; Score 107.6; DB 9; Length 496;
Best Local Similarity 78.0%; Pred. No. 2.4e-12;
Matches 142; Conservative 0; Mismatches 39; Indels 1; Gaps 1;

QY 208 TTCTGTTCTTCTTATTTGTCATATTTTATTTGTTTACTGATATATACGAGTAGTTAT 267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 429 TTTTTCCTTCGATTTGGCAAT-GTTTCATTTGCTGTTACTGCTATATGCGAGTAGTTAT 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 268 ACATACATGCATACATATCATATCATATATATATATATATATATATATATATATAT 327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 GCACATGCATACATATGTCATCATATATATATATATATATATATATATATATAT 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 328 CTAATAAGTCAATATTTCTACACACGACATGTTATGTTTCTCCACAACTTTAC 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 CTAATAAGTCAATATTTCTACACAAATACTAAAGTTTGTGGTCCAAACATCAATAC 251

QY 388 CT 389
|
Db 250 AT 249

RESULT 4
BI885732
LOCUS
DEFINITION
5' similar to contains element MSRI repetitive element ;, mRNA
sequence.
ACCESSION
BI885732
VERSION
BI885732.1 GI:16093003
KEYWORDS
zebrafish.
SOURCE
Danio rerio
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 572)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Fape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ReissourcenzentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 366.
Location/Qualifiers
1..572
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="4490204"
/clone_lib="zebrafish WashU MPING EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XL1-blue MRF"

FEATURES
source

```

/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer (5'pgactagttctagatcgagcgccgctttttttttttttttt3'). double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

```

BASE COUNT      13 a      421 c      44 g      16 t      78 others
ORIGIN

Query Match      4.4%; Score 58.2; DB 10; Length 572;
Best Local Similarity 43.6%; Pred. No. 0.048;
Matches 159; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 702 CCGCCTTTATTTTCAGGCTCGCTGGCCCCACGGCGCTGCTCGGTGCACGAGGCGACTACC 761
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 CCCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 135
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 762 CCAACCTCTCACGAAACCCGCGCTGGATCGGCAATCAACGAGGTGTCGCCGCTGC 821
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 CCCCCCCCCCCCCCCCCCCCCCGCGCGGCCCCCCCCCCCCCCCCCCCCCCCCCCCC 195
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 822 CCACTCTCCAGCTCCACGGCACCATCCCTCTGTCAGCGCTGCACGAGCATGCGTGTGC 881
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 255
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 882 GGAACGGCACACACCCCAACCCCACTCAGAAACCCGCTGCCGCGCTGCCGCTGTGC 941
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 315
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 942 GTCCGCGCTCGGCAACGAGCGCGCGCGCTGCTAGTCTCCCTGGAGACACCGACACCTG 1001
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 375
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1002 TGGGCCCTTTGTTTATTCATCCGAAATCTCATCTGCCGCCACGCGCGCTGCCGCTGC 1061
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CCCCCCCTCTGTTGTTTTCNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 435
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1062 CGCCC 1066
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 CCCCC 440
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 5
CNS0106X/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN03k20 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL098595
VERSION
AL098595.1 GI:5610206
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)

```

COMMENT

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> -. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CDPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Burcheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

```

FEATURES
  source
    Location/Qualifiers
      1..1101
      /organism="Drosophila melanogaster"
      /plasmid="pBeloBAC11"
      /db_xref="taxon:7227"
      /clone_lib="DrosBAC"
      /clone="BACN03K20"
      /note="end : 77"
BASE COUNT      258 a      107 c      60 g      175 t      501 others
ORIGIN

```

Query Match 4.4%; Score 58.2; DB 12; Length 1101;  
Best Local Similarity 19.2%; Pred. No. 0.049;  
Matches 87; Conservative 182; Mismatches 184; Indels 0; Gaps 0;

[illegible]

RESULT 6					
AU087781/c	AU087781	500 bp	linear	EST 27-JAN-2001	
LOCUS	AU087781	Sugano	Malaria cDNA library	Plasmodium falciparum	CDNA
DEFINITION	clone xPfn5810,	mrna	sequence.		

ACCESSION AU087781  
VERSION AU087781.1 GI:12389922  
KEYWORDS EST.  
SOURCE malaria parasite *P. falciparum*.  
ORGANISM Plasmodium falciparum  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 500)  
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.  
TITLE FULL-malaria: a database for a full-length enriched cDNA library  
from human malaria parasite, Plasmodium falciparum  
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)

MEDLINE	COMMENT

20574754  
Contact: Junichi Watanabe  
Institute of Medical Science  
The University of Tokyo, Department of Parasitology  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Tel: 81-3-5449-5378  
Fax: 81-3-5449-5410  
Email: jwatanabe@nagae.ims.u-tokyo.ac.jp  
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Suganono, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* 200 (1-2), 149-156 (1997).

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital Status</i>	3. <i>Marital Status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Income</i>	5. <i>Income</i>
6. <i>Occupation</i>	6. <i>Occupation</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political Affiliation</i>	8. <i>Political Affiliation</i>
9. <i>Health Status</i>	9. <i>Health Status</i>
10. <i>Travel History</i>	10. <i>Travel History</i>
11. <i>Employment Status</i>	11. <i>Employment Status</i>
12. <i>Family Size</i>	12. <i>Family Size</i>
13. <i>Home Ownership</i>	13. <i>Home Ownership</i>
14. <i>Vehicle Ownership</i>	14. <i>Vehicle Ownership</i>
15. <i>Insurance Status</i>	15. <i>Insurance Status</i>
16. <i>Charitable Contributions</i>	16. <i>Charitable Contributions</i>
17. <i>Volunteer Work</i>	17. <i>Volunteer Work</i>
18. <i>Substance Use</i>	18. <i>Substance Use</i>
19. <i>Mental Health</i>	19. <i>Mental Health</i>
20. <i>Life Satisfaction</i>	20. <i>Life Satisfaction</i>

```

Location/Qualifiers
1..500
/organism="plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/clone_lib="XpFn5810"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
228 a 25 c 33 g 203 t 11 others

```

BASE COUNT	228 a	25 c	33 g	203 t	11 others
ORIGIN					

Query Match 4.4%; Score 57.8; DB 9; Length 500;  
Best Local Similarity 54.2%; Pred. No. 0.058;

QY 220 TATTGTCATATTTTATTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279

224 TATTGTCATATTTTATTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 215

D0  
Z74  
Q07

1ATAAATGTAAGTAAATTCAGCAATTTCTGCTAAAATTAAAACTAAAAATGACT 339

[illegible]

340 AAATTCTAACCCACGACATTGTAATGTTTTCTCCACAACATTACCTATTCTACATT 399

Db

154 TATTAT-TAAATTTTAACTTAATGTGCTTCFPAATAAFAATTANATAAA 96

400 GTTCTATTTTCGAATTTTCACTCTATAAACAACATAGTCTACAATGGAAAAACAGTGCCTTGT 459

Db  
95 TTTTCTTTTATATAAATNCGACGAAGTGAAAAATGACTATTGTAAAAAAATATTTCTGGAT 36

QY 460 ACGACTATATA 470

Db 35 ACAATAAATACA 25

## RESULT 7

CNS006XK  
935 bp  
DNA  
linear  
GSS 03-JUN-1999

LOCUS  
CNS006XK/C  
DEFINITION  
*Brosophylla melanogaster* genome survey sequence T7 end of BAC #

BAC14N09 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.

ACCESSION	AL066051
VERSION	AL066051.1
	GI:4945019

**KEYWORDS**  
GSS.  
fruit fly.  
**SOURCE**

ORGANISM *Drosophila melanogaster*  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS  
1 (bases 1 to 935)  
Genoscope.  
Sequence

JOURNAL  
TITLE  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
Direct Submission  
Genoscope - FRANCE (E-mail : searef@genoscope.cns.fr)  
Genoscope - FRANCE (E-mail : searef@genoscope.cns.fr)

BP 191 31000 EVRY cedex  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a

COMMENT  
Determination of the gene order and  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila

the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and



```
FEATURES
Source
Location/Qualifiers
1..1058
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN1117"
/note="end : SP6"

BASE COUNT      121 a   175 c   93 g   396 t   273 others
ORIGIN
Query Match      4.2%; Score 55.6; DB 12; Length 1058;
Best Local Similarity 32.9%; Pred. No. 0.17;
Matches 78; Conservative 65; Mismatches 94; Indels 0; Gaps 0;

QY 186 ATCGGTTACTCCGATAGGACATTCGTTCTTCTTATTTGTCATATTTTATTTGTT 245
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 700 ATAGSCWATTCWSASTRATASWMTSTSWTBWBVKRSTRATKAKTBATCTVTTWTATA 759
QY 246 TACTGATTATACGAGTAGTATATACATACATGACATATCATCATATATACCAAT 305
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 760 HSMTSWYTTACWTSMTSWSWTACACWCYCTACCCAWTATATCTMTATATHCACATW 819
QY 306 ATTTTCTTAATTAATTAATAAAGTAAATGACTAAATTTCTAACACACGACATTTGA 365
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 820 ACATTSCYTMATPAAHTTCACTWAANAAYATTSTAWATYMTTSTCTTSWVSAAYAA 879
QY 366 ATGTTTCTCCACACACTTACCTATCTACATGTTCTTATTCGAATTTCACTCTA 422
||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 880 TASWTTBTACSMPTACTSAYASTHTAMTATATATATATATATATATATATATATAT 936

RESULT 10
BI952390/c      818 bp      mRNA      linear      EST 19-OCT-2001
LOCUS
DEFINITION
HVSME0006C23f Hordeum vulgare green seedling EST library
HVCNDA0014 (Blumeria infected) Hordeum vulgare cDNA clone
HVSME0006C23f, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hordeum vulgare
Barley.
REFERENCE
AUTHORS
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons
J., Oates, R. and Main, D.
TITLE
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
JOURNAL
COMMENT
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 224
Seq primer: AATTAAACCTCACTAAAGG
High quality sequence stop: 808.

FEATURES
Source
Location/Qualifiers
1..818
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSME0006C23f"
/clone_lib="Hordeum vulgare green seedling EST library
HVCNDA0014 (Blumeria infected)"

/tissue_type="green seedling leaf"
/lab_host="TJC121"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders
Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT      74 a   142 c   556 g   46 t

Query Match      4.2%; Score 55.2; DB 10; Length 818;
Best Local Similarity 46.4%; Pred. No. 0.21;
Matches 180; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 702 CGCGCTTTATTTTCAGGCTCGTGGCCACGGGGTGTGTCGACGAAGGCACCTACC 761
Db 635 CCCCCTCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 566
QY 762 CCAACTCTACCGAAAAACCGCGTGGATCGGCAATCAACAGAGTGTGTCGCCGTGC 821
Db 565 CCCCCTCTCCCGCCCCCTCCCTCCCTCCCGCCCCCCCCCCCCCCCCCCCCCCCC 506
QY 822 CCACTCTCCAGTCCACGGACCATCTCTGACGCCCTCACCAGCATCGCGTGTGC 881
Db 505 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 446
QY 882 GGAACGGCACAACACACCCCAACCACTCAGAAACCCCGTCCCGCGGTGCGTGTGC 941
Db 445 CCCCCCCCCCTCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCT 386
QY 942 GTCCGCGTCCGCAACGAGGGCGCGCGTGTGAGTCCCTGGACACCGACACCTGTG 1001
Db 385 TCCCGCCCCCTCAGCGCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 326
QY 1002 TCGGCGCTTGTATTATTCATCCGAAATCTCATCTGCCCGCCCGCCGCTGCGC 1061
Db 325 CCCCCCCCCCGCGCATCTCCCGCCCCCGCCGACCTGCGCCGCTCCCGCCGCGCC 266
QY 1062 CGCCCGGATATATATACCATCGTTATC 1089
Db 265 CGCGCGCACAGCCCCCGCTTCTCCATC 238

RESULT 11
CNS0091P
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR19p16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
AL053013
AL053013.1 GI:4934461
```





VERSION	AQ251130.1	GI:3704196	
KEYWORDS	GSS.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 798)		
REFERENCE	Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.		
AUTHORS	BAC End Sequences at ATGC		
TITLE	Unpublished (1997)		
JOURNAL	Contact: Ecker J.		
COMMENT	Arabidopsis Thaliana Genome Center University of Pennsylvania Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104 Tel: 215-898-9384 Fax: 215-898-8780 Email: jecker@atgenome.bio.upenn.edu Seq primer: Sp6 Class: BAC ends.		
FEATURES	Location/Qualifiers		
source	1..798		
	/organism="Arabidopsis thaliana"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	/clone="F26H6"		
	/clone_lib="IGF"		
	/sex="hermaphrodite"		
	/note="Vector: BelobAC11; Site_1: ECORI; Site_2: ECORI; Produced by Thomas Altmann"		
BASE COUNT	347 a	99 c	259 t
ORIGIN	3 others		
Query Match	4.1%;	Score 54;	DB 12; Length 798;
Best Local Similarity	53.9%;	Pred. No. 0.37;	
Matches 132;	Conservative	0;	Mismatches 111; Indels 2; Gaps 1
QY	226	TGCATATTTTTTATTGTTGTTACTGATTATACGAGTAGTTATACATACATGCACATACAT	285
Db	152	TGATTATGCTATAGTATCAAAAGGTGAATTAATAGTATTAAACATATAATAATTAGTAA	211
QY	286	ATCATCATATATCACAAATATTTTCTAAATTAATTAACATGCTAAATTTT	345
Db	212	AAATCATATTTTGAATTTTGTATTATTAATTTCAATTTAAATTTAAAGTTCTGCATAAT	271
QY	346	CTAACACCAACGACATGTAATGTTTCTCCAAACACTTTACCTATTCTACATGTTCTA	405
Db	272	NTAATAAATATATATATAAATTCCTCCAAACATTTATTTATTTTATAATAAAAA	331
QY	406	TTTCGAATTTCACTCTATAAACACATAGTCTACAATGGAAACAGTCTTTGTACGACT	465
Db	332	TT--AAATATATTACAAATTTAAATAAGTTTCACTTTAAACATTTATTTATTTATACTAAT	389
QY	466	ATATA	470
Db	390	TTAAA	394
RESULT 14			
CNS00FVE			
LOCUS	CNS00FVE	1101 bp	DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR32P18 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL071298		
VERSION	AL071298.1	GI:4951138	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		

Search completed: November 7, 2002, 12:25:20  
Job time : 1257.97 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:47:38 ; Search time 1685.86 Seconds  
(without alignments)  
16248.621 Million cell updates/sec

Title: US-09-905-558C-16

Perfect score: 1309

Sequence: 1 cccatcgctgcttctgtac.....caagggaagtgatcccatg 1309

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_to.\*

11: gb\_ats.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_ats.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
-----					

1	232.8	17.8	2126	6	AX299951
2	124.6	9.5	1888	6	AX099709
3	112.2	8.6	9285	8	AX099709
4	60.2	4.6	174802	2	ZM009989
5	58	4.4	125020	9	AF429315
6	56.6	4.3	173848	9	AF429315
7	56.2	4.3	169660	9	AF429315
8	55.6	4.2	204652	2	AF429315
9	55.4	4.2	65523	2	AF429315
10	55.2	4.2	650	3	AF429315
11	54.8	4.2	204652	2	AF429315
12	54	4.1	89904	8	AF429315
13	54	4.1	196286	8	AF429315
14	53.8	4.1	115522	8	AF429315
15	53.6	4.1	214033	2	AF429315
16	53.4	4.1	614	3	AF429315
17	53.2	4.1	167553	9	AF429315
18	53	4.0	127539	9	AF429315
19	53	4.0	146868	2	AF429315
20	53	4.0	151117	2	AF429315
21	53	4.0	161277	2	AF429315
22	53	4.0	176033	2	AF429315
23	53	4.0	178342	9	AF429315
24	53	4.0	187150	2	AF429315
25	53	4.0	199385	2	AF429315
26	52.6	4.0	72090	2	AF429315
27	52.6	4.0	80692	2	AF429315
28	52.6	4.0	110000	2	AF429315
29	52.6	4.0	141532	2	AF429315
30	52.6	4.0	191772	2	AF429315
31	52.6	4.0	293431	2	AF429315
32	52.4	4.0	171574	2	AF429315
33	52.4	4.0	196203	9	AF429315
34	52.2	4.0	90487	9	AF429315
35	52.2	4.0	136801	9	AF429315
36	52.2	4.0	149450	2	AF429315
37	52.2	4.0	153300	9	AF429315
38	52.2	4.0	170141	2	AF429315
39	52.2	4.0	213692	2	AF429315
40	52	4.0	159593	9	AF429315
41	51.8	4.0	678	3	AF429315
42	51.8	4.0	14135	3	AF429315
43	51.6	3.9	79646	8	AF429315
44	51.6	3.9	162996	8	AF429315
45	51.4	3.9	64148	2	AF429315

## ALIGNMENTS

RESULT 1  
AX299951  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
BASE COUNT  
ORIGIN

AX299951  
Sequence  
AX299951  
AX299951.1  
GI:17129442  
Zea mays.  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade: Panicoideae; Andropogoneae; Zea.  
1 (sites)  
Conner, T.W., Dubois, P., Malyen, M. and Masucci, J.D.  
Plant regulatory sequences for selective control of gene expression  
Patent: WO 0183790-A 92 08-NOV-2001;  
Monsanto Technology LLC (US)  
Location/Qualifiers  
1. .2126  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
607 a 419 c 433 g 667 t

AX299951  
Sequence  
2126 bp  
DNA  
linear  
PAT 26-NOV-2001  
Sequence 92 from Patent WO0183790.  
AX299951.1  
GI:17129442  
Zea mays.  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade: Panicoideae; Andropogoneae; Zea.  
1 (sites)  
Conner, T.W., Dubois, P., Malyen, M. and Masucci, J.D.  
Plant regulatory sequences for selective control of gene expression  
Patent: WO 0183790-A 92 08-NOV-2001;  
Monsanto Technology LLC (US)  
Location/Qualifiers  
1. .2126  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
607 a 419 c 433 g 667 t

```
Query Match 17.8%; Score 232.8; DB 6; Length 2126;
Best Local Similarity 83.0%; Pred. No. 2.7e-44;
Matches 289; Conservative 0; Mismatches 57; Indels 2; Gaps 2;

QY 5 TCGCTGCTTTGTACATCATGTTCTTCATCATCTCCCGCAGCGGCGTGTGCTGT 64
DB 82 TCGGCTGATGTCACAGCTGCTGTCGTCGCCCTCCCAAGTTGACGGTGTGCTGT 141
QY 65 CTTATTCA-GACTACCGTTCGAGTGACTGCGATGCGGTACATCTTTCTGATCGACTTTGT 123
DB 142 CTTCTCCCGGCGACCGTTCGAGGAGTGCACCTGCGTACATCTTCCTGCCACGACTTCGT 201
QY 124 ACGGCTACATCGACATATACACAGAGTGTCTGTCGTGTAATGAGTCACTAATGCGCTTAA 183
DB 202 ACGGCTACATCGAACAACACAGAGATGTCCTGTCGTGTAATGAGTCACTAATGCGCTTGA 261
QY 184 GCATCGGTACTCGGTAGGTACATCTCTTCTTCTTCTTATTTGTGTCATATTTTATTGTG 243
DB 262 GCATCGGTCTCGCTGGGTACACTGTCTCTGTAATTTGTGCAAT-GTTTCATTGCTG 320
QY 244 TTTACTGATATACGAGTAGTTATATACATACATGACATATATATATATATATATATAC 303
DB 321 TTTACTGATATACGAGTAGTTATATACATATGACATATATATATATATATATATATAT 380
QY 304 ATATTTTCTAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 351
DB 381 CTGATTATCTGGATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 428

RESULT 2
AX099709/c
LOCUS AX099709 1888 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 65 from Patent WO0119976.
ACCESSION AX099709
VERSION AX099709.1 GI:13538763
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 1888)
Anderson,H.M., Chay,C.A., Chen,G. and Conner,T.W.
Plant regulatory sequences for control of gene expression
Patent: WO 0119976-A 65 22-MAR-2001;
MONSANTO COMPANY (US)
FEATURES
Location/Qualifiers
source 1..1888
/organism="Zea mays"
/db_xref="taxon:4577"
BASE COUNT 569 a 443 c 423 g 453 t
ORIGIN
Query Match 9.5%; Score 124.6; DB 6; Length 1888;
Best Local Similarity 63.9%; Pred. No. 8.4e-19;
Matches 227; Conservative 0; Mismatches 114; Indels 14; Gaps 2;

QY 78 CCGTTCGAGTACTCGAGCGGTACATCTTTCTGATCGACTTTGTACGGCTACATCGAA 137
DB 784 CTGTTTAGGGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 725
QY 138 CATATACAGGAGATGCTCGTGTGAATAGTACTAATAGCTTAAGCATCGGTACTTCC 197
DB 724 CATACACAGGAGATGCTGTTGTGAATTTGGCGCGTGTACTTTTGAGCATCGGTCCCTC 665
QY 198 GTAGGATACATCTGTTCTTCTTATTTTGGCATATTTTATTTGTTGTTTACTGATTATAC 257
DB 664 ACAGGAAGTCTCTGTTTCCATCTTTT--TGGGTGCTCTTTTGTGCTGCTTATATAC 607
QY 258 GAGTAGTATATACATATGATATATATATATATATATATATATATATATATATATATAT 317
DB 606 GAATAGTGT-----ATACATGCTGCTACATATATATATATATATATATATATATAT 559
```

```
QY 318 TAAATTAACACTAAACATGACTTAAATTTCTTAACACACACAGCATTTGTAATGTTTCTCCA 377
DB 558 TAAATTAACACTAGTAATGCTTCATTTTCAACATATTTACAGTAATAAAGTTAGGTA 499
QY 378 ACAACTTTACCTATTCTACATTTGTTCTATTTCGAATTTCACTCTATATAACACACAT 432
DB 498 TCAGCTAATTTCTTGTGATAGGTATAAATGACGATGCTCTATATATTTATCAGAT 444

RESULT 3
ZM009989/c
LOCUS ZM009989 9285 bp DNA linear PLN 06-DEC-2000
DEFINITION Zea mays D3L H(+)-transporting ATPase (Mhal) gene, complete cds.
ACCESSION U09989
VERSION U09989.1 GI:507770
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 2391 to 2453; 2525 to 2659; 2751 to 2930; 3019 to 3225;
3306 to 3425; 3510 to 3629; 3722 to 3844; 3927 to 3938)
Jin,Y. and Bennetzen,J.L.
Integration and Nonrandom Mutation of a Plasma Membrane Proton
ATPase Gene Fragment within the Bsl Retroelement of Maize
Plant Cell 6, 1177-1186 (1994)
95003707
REFERENCE
2 (bases 1 to 9285)
Jin,Y.
Direct Submission
Submitted (24-MAY-1994) Young-Kwan Jin, Biological Sciences, Purdue
University, 339 Hansen Life Science Research Building, West
Lafayette, IN 47907, USA
FEATURES
Location/Qualifiers
source 1..9285
/organism="Zea mays"
/strain="D3L"
/db_xref="taxon:4577"
/chromosome="2L"
/map="2L"
/clone="Cl8"
/clone_lib="Sau3AI partial of maize genomic DNA into
Charon 40"
join(<1446..1508,1874..1993,2105..2203,2316..2453,
2525..2659,2751..2930,3019..3225,3306..3425,3510..3629,
3722..3844,3927..4031,4113..4257,4379..4543,4637..4875,
4964..4996,5091..5172,5257..5417,5513..5689,5768..5950,
6036..>6290)
/gene="Mhal"
/product="H(+)-transporting ATPase"
join(<1446..1508,1874..1993,2105..2203,2316..2453,
2525..2659,2751..2930,3019..3225,3306..3425,3510..3629,
3722..3844,3927..4031,4113..4257,4379..4543,4637..4875,
4964..4996,5091..5172,5257..5417,5513..5689,5768..5950,
6036..>6290)
/gene="Mhal"
/codon_start=1
/number=1
join(1446..1508,1874..1993,2105..2203,2316..2453,
2525..2659,2751..2930,3019..3225,3306..3425,3510..3629,
3722..3844,3927..4031,4113..4257,4379..4543,4637..4875,
4964..4996,5091..5172,5257..5417,5513..5689,5768..5950,
6036..>6290)
/gene="Mhal"
/codon_start=1
/product="H(+)-transporting ATPase"
/protein_id="AAB60276.1"
/db_xref="GI:533775"
/translation="MGTTEDKASNDLAVLKEAVDLENIAIOEVFESLRCSPOGLSTEQ
AEORLAIFGPNKLEEKQESKFLKFLGMNPLSWWMAAAIMAILANGNRPDPWDQ
```







TITLE A repeat expansion in the gene encoding juncctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES

source Location/Qualifiers

1. .125020

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S20 and WI-12410"

/note="Isolated from a patient with Huntington's Disease-Like 2 (HDL2)"

repeat\_region complement(35581..35746)

/rpt\_type=tandem

/rpt\_unit=ctg

complement(<36507..>36887)

/gene="JPH3"

/note="JPH3"

complement(<36507..36887)

/gene="JPH3"

/note="JPH3"

complement(<36507..36887)

/gene="JPH3"

/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"

/codon\_start=1

/product="juncctophilin 3"

/protein\_id="AAL40941.1"

/db\_xref="GI:17648245"

/translation="MSGGRNFDDGSGYCGWEDGKAHGHVCTGPKGGEYTGWS HGFVLGYTPSGNTYQGTWQAQGRHGLGSKWYKGEWTHGFKRYGVRECA GNAKYEGTWSNGLQDGYGTETYSQ"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.4%; Score 58; DB 9; Length 125020;

Best Local Similarity 11.8%; Pred. No. 0.01;

Matches 70; Conservative 248; Mismatches 273; Indels 0; Gaps 0;

QY 699 GGGCCGCTTTATTTTCAGCTCGCTGGCCCGGCGGCGTGCAGGAGGCACT 758

Db 16847 GGGACCCAGTAGTAGCGCCATCGCARRKMKSHAGRRMCIYKSKWMSBMSVSYVK 16906

QY 759 ACCCCNACTCTCACGGAACCGCGCTGGATCGGCAATCAACGAGGTGGTCCCGGT 818

Db 16907 MHSMHASBSCHWBKMTWSSCCSMNMYKSSWWSGMCWRRRSKWKWYSGMSRSK 16966

QY 819 GCCACTCTCCAGTCCAGCGGACCATCCCTCTGACGGCTGCACGCGATGCGGTGC 878

Db 16967 SMRYTGGKMKRSSMWTSSCYASCMCCWCCMCCMCRSCCYCCMYCCACKCYM 17026

QY 879 GCGGAACGCGCACACACCCCAACCCACTCACGAAACCCGCTCCGCGGCGCGGTGT 938

Db 17027 SSVYWTASYSMSRRSYWRSMKCSWBSRSCSKSRGCGGSGKSGKSYGRKTKR 17086

QY 939 CGGTCCGCGCTCGGAACGAGCGCGCGCTGTGATGCCCTGGACACCCGACACC 998

Db 17087 KSRGKHGAKWMTYRRSRMRKMKYSSRGMYCWCWGRRCYCSMTSRAMCCSYCY 17146

QY 999 TGTGCGCCCTTGTATTATTCATCCGAATCTCATCTGCCCGCCGCGGCGGCTGTC 1058

Db 17147 AKCKSNVCYCYGSMKGYIYACSTRGSMKCYMRGTYSTSCGCCCTTTTCCCCNA 17206

QY 1059 GCCCGCCCGGATATATACCCATCGTTATTCATCGATCGATCGGCTCACTCACGGTAGC 1118

Db 17207 NTGGGAAGCTTTTNCNKKTSYKRNKGCAMCKYNNYNNWSRSSCRAGMSCTKYKSSM 17266

QY 1119 TCATGCTCAGCGCTACGATCGAGAACTTATTTCGCTGCTGCCAGTCTCCGCTCCG 1178

Db 17267 TWSMASYCWMSMYCYMSRSMASRGMSWSSYMKMKMSKSRMSYMKCCWCMKCYCMRS 17326

QY 1179 GTGCCCTTCAGTCTGTCTCACACTAGCTGCTGTGGACACATCGAAGTGGTGTGTCAGCT 1238

Db 17327 MRSRGSYTYWASWSSSSRGCTCTRCYCMSSKCYKYKSYMMRSMKMGKMSRWGSM 17386

QY 1239 AGCTAGCTCGCGGTGACACGACATGACCGCAGTGTGCGCGGGGCTGAT 1289

Db 17387 WGSASRSSCYKCYKSMRCSMSSSKYRCACGMMKGGYMYRCHWSMKRRW 17437

RESULT 6

AC090043 173848 bp DNA linear PRI 11-FEB-2001

LOCUS Homo sapiens chromosome 3 clone RP11-551L4 map 3p, complete

DEFINITION sequence.

AC090043

VERSION AC090043.1 GI:12745081

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 173848)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H.F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.

TITLE Chromosome 3p genomic sequence

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 173848)

Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Luo, Y., Ni, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.

Direct Submission

Submitted (11-FEB-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China

Center: Beijing Center

Center code: Beijing

Website: http://hg.cigtp.ac.cn

http://www.genomics.org.cn

Contact: hgc@igtp.ac.cn

Project Information

Center project name: 1# project

Center clone name: RP11-551L4

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator; ET 55% of reads

Assembly: Dye-terminator Big Dye; 45% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 501 bases at least Q40

Consensus quality: 829 bases at least Q30

Consensus quality: 1046 bases at least Q20

Insert size: 1157; sum-of-contigs

Quality coverage: 1.35x in Q20 bases; sum-of-contigs

Location/Qualifiers

1. .173848

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

FEATURES

source

Query Match

Best Local Similarity

Matches 208;

4.3%;

Score 56.6;

DB 9;

Length 173848;

Indels 1;

Gaps 1;

BASE COUNT

ORIGIN

56663 a

31451 c

31882 g

53852 t

map="3p"

clone="RP11-551L4"

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr9

This sequence is the entire insert of clone RP11-382H24. The true left end of clone RP11-664D14 is at 135950 in this sequence. The true right end of clone RP11-187K14 is at 53851 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-382H24 is from the library RPI-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6.

FEATURES

Source

1. 169660

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="9"

/map="p22.1-23"

/clone="RP11-382H24"

/clone\_lib="RPI-11.2"

2. 492

/note="match: GSS: Em:AQ700717"

/complement(1479..2009)

/note="match: GSS: Em:AQ569033"

1589..1778

/note="MIR repeat: matches 23. 192 of consensus"

1669..1724

/note="2 copies 28 mer 92% conserved"

2068..2401

/note="LMC4 repeat: matches 7453. 7834 of consensus"

2409..2828

/note="15 copies 28 mer 62% conserved"

2429..2848

/note="7 copies 60 mer 65% conserved"

2452..2801

/note="175 copies 2 mer ta 61% conserved"

2819..2911

/note="LMC4 repeat: matches 7864. 7961 of consensus"

/complement(4102..4478)

/note="match: GSS: Em:AQ180899"

4497..5002

/note="match: GSS: Em:AQ716496"

4621..4865

/note="MIR repeat: matches 1. 261 of consensus"

5200..5294

/note="L2 repeat: matches 2578. 2670 of consensus"

6031..6654

/note="LIPAL2 repeat: matches 5583. 6152 of consensus"

7419..7966

/note="match: GSS: Em:AQ378047"

7431..7737

/note="ALUSC repeat: matches 1. 307 of consensus"

7763..8103

/note="MLT1A1 repeat: matches 205. 536 of consensus"

8379..8422

/note="22 copies 2 mer tg 86% conserved"

9171..9447

/note="LM4 repeat: matches 2900. 3176 of consensus"

9450..9739

/note="LM4 repeat: matches 3869. 4175 of consensus"

9774..9835

/note="LIM47 repeat: matches 6223. 6284 of consensus"

10099..10533

/note="LM47 repeat: matches 6223. 6284 of consensus"

repeat\_region repeat: matches 9. 466 of consensus"

RESULT 7

AL161449

LOCUS

DEFINITION

169660 bp

DNA

linear

PRI 04-DEC-2001

Human DNA sequence from clone RP11-382H24 on chromosome 9p22.1-23

Contains a thioredoxin peroxidase pseudogene, a SSB (Sjogren syndrome antigen B (autoantigen Ia) pseudogene, the 3' end of the MPDZ gene for multiple PDZ domain protein and a CpG island, complete sequence.

AL161449

AL161449.7

GI:8894259

HTG; CpG island; MPDZ.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 169660)

Sehra.H.

Direct Submission

Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Jul 1, 2000 this sequence version replaced gi:8653807.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information



correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

\* NOTE: This is a 'working draft' sequence.

\* as soon as it is available and the accession number will

\* be preserved.

#### FEATURES

source  
Location/Qualifiers  
1. 204652  
/organism="Plasmodium falciparum"  
/strain="3D7"  
/db\_xref="taxon:5833"  
/chromosome="13"

BASE COUNT 68880 a 17396 c 17710 g 67856 t 32810 others  
ORIGIN

Query Match 4.2%; Score 55.6; DB 2; Length 204652;  
Best Local Similarity 52.6%; Pred. No. 0.041;  
Matches 121; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
Qy 204 TACATTCGTTCTTCTTATTTGTCATATTTTATTGTTTCTACTGATATACGAGTAG 263  
Db 27538 TTCTATGTTGTCATATTTGTCATATTTTATTGTTTCTACTGATATATATAATAT 27597  
Qy 264 TTATACATACATGCACATACATATCATACATATATCATACATATTTTCTAAATTAAT 323  
Db 27598 ATATACATATACACATACATGCATACATATATATATATATATATATATATATTA 27657  
Qy 324 AAACTAAAATGACTAAATTTCTAACACCAACGACATTTGTAATGTTTCTCCAACT 383  
Db 27658 TAACAAAAAATAAATAATATTTATTTCCCAAAATAATATATGATATATATATTA 27717  
Qy 384 TTACTATTCATATGTTTCTATTTTCTGAATTTCTACTCTATTAACAAACATA 433  
Db 27718 TGTCTAAATAGTTTTTTCATATATATATATATATATATATATATATATTAACATA 27767

RESULT 9  
AC024266/c 65523 bp DNA linear HTG 13-JUL-2000  
LOCUS Homo sapiens clone RP11-115F10, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC024266  
ACCESSION AC024266.1 GI:7107792  
VERSION  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 65523)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-115F10

Unpublished

2 (bases 1 to 65523)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,

Bouckhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,

Chapel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,

Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M.,

Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,

Galan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,

Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R.,

Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,

McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,

Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M.,

Peterson, K., Plierre, N., Pisanic, C., Pollara, V., Raymond, C.,

Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

#### TITLE

JOURNAL

#### COMMENT

Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,  
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and  
Zody, M.  
Direct Submission  
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: LJ997

Center clone name: 115\_F\_10

\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved. 793: contig of 793 bp in length

1 794 893: gap of 100 bp

894 1687: contig of 794 bp in length

1688 1787: gap of 100 bp

1788 2572: contig of 785 bp in length

2573 2672: gap of 100 bp

2673 3486: contig of 814 bp in length

3487 3586: gap of 100 bp

3587 4391: contig of 805 bp in length

4392 4491: gap of 100 bp

4492 5298: contig of 807 bp in length

5299 5398: gap of 100 bp

5399 6142: contig of 744 bp in length

6143 6242: gap of 100 bp

6243 7033: contig of 791 bp in length

7034 7133: gap of 100 bp

7134 7941: contig of 808 bp in length

7942 8041: gap of 100 bp

8042 8832: contig of 791 bp in length

8833 8932: gap of 100 bp

8933 9736: contig of 804 bp in length

9737 9836: gap of 100 bp

9837 10619: contig of 783 bp in length

10620 10719: gap of 100 bp

10720 11503: contig of 784 bp in length

11504 11603: gap of 100 bp

11604 12326: contig of 723 bp in length

12327 12426: gap of 100 bp

12427 13201: contig of 775 bp in length

13202 13301: gap of 100 bp

13302 14085: contig of 784 bp in length

14086 14185: gap of 100 bp

14186 14981: contig of 796 bp in length

14982 15081: gap of 100 bp

15082 15879: contig of 798 bp in length

15880 15979: gap of 100 bp

15881 16745: contig of 766 bp in length

16746 16845: gap of 100 bp

16846 17645: contig of 800 bp in length

17646 17745: gap of 100 bp

17746 18541: contig of 796 bp in length

18542 18641: gap of 100 bp

18642 19447: contig of 806 bp in length

```
* 19448 19547: gap of 100 bp
* 19548 20376: contig of 829 bp in length
* 20377 20476: gap of 100 bp
* 20477 21178: contig of 703 bp in length
* 21180 21279: gap of 100 bp
* 21280 22064: contig of 785 bp in length
* 22065 22164: gap of 100 bp
* 22165 22935: contig of 771 bp in length
* 22936 23035: gap of 100 bp
* 23036 23851: contig of 816 bp in length
* 23852 23951: gap of 100 bp
* 23952 24827: contig of 876 bp in length
* 24828 24927: gap of 100 bp
* 24928 25698: contig of 771 bp in length
* 25699 25798: gap of 100 bp
* 25799 26586: contig of 788 bp in length
* 26587 26686: gap of 100 bp
* 26687 27505: contig of 819 bp in length
* 27506 27605: gap of 100 bp
* 27606 28408: contig of 803 bp in length
* 28409 28508: gap of 100 bp
* 28509 29294: contig of 786 bp in length
* 29295 29394: gap of 100 bp
* 29395 30181: contig of 787 bp in length
* 30182 30281: gap of 100 bp
* 30282 31028: contig of 747 bp in length
* 31029 31128: gap of 100 bp
* 31129 31928: contig of 800 bp in length
* 31929 32028: gap of 100 bp
* 32029 32815: contig of 787 bp in length
* 32816 32915: gap of 100 bp
* 32916 33724: contig of 809 bp in length
* 33725 33824: gap of 100 bp
* 33825 34585: contig of 761 bp in length
* 34586 34685: gap of 100 bp
* 34686 35469: contig of 784 bp in length
* 35470 35569: gap of 100 bp
* 35570 36370: contig of 801 bp in length
* 36371 36470: gap of 100 bp
* 36471 37277: contig of 807 bp in length
* 37278 37377: gap of 100 bp
* 37378 38157: contig of 780 bp in length
* 38158 38257: gap of 100 bp
* 38258 39015: contig of 758 bp in length
* 39016 39115: gap of 100 bp
* 39116 39853: contig of 738 bp in length
* 39854 39953: gap of 100 bp
* 39954 40734: contig of 781 bp in length
* 40735 40834: gap of 100 bp
* 40835 41594: contig of 760 bp in length
* 41595 41694: gap of 100 bp
* 41695 42497: contig of 803 bp in length
* 42498 42597: gap of 100 bp
* 42598 43390: contig of 793 bp in length
* 43391 43490: gap of 100 bp
* 43491 44319: contig of 829 bp in length
* 44320 44419: gap of 100 bp
* 44420 45162: contig of 743 bp in length
* 45163 45262: gap of 100 bp
* 45263 46047: contig of 785 bp in length
* 46048 46147: gap of 100 bp
* 46148 46913: contig of 766 bp in length
* 46914 47013: gap of 100 bp
* 47014 47792: contig of 779 bp in length
* 47793 47892: gap of 100 bp
* 47893 48669: contig of 777 bp in length
* 48670 48769: gap of 100 bp
* 48770 49560: contig of 791 bp in length
* 49561 49660: gap of 100 bp
* 49661 50399: contig of 739 bp in length
* 50400 50499: gap of 100 bp
* 50500 51301: contig of 802 bp in length
* 51302 51401: gap of 100 bp
```

```
* 51402 52227: contig of 826 bp in length
* 52228 52327: gap of 100 bp
* 52328 53120: contig of 793 bp in length
* 53121 53220: gap of 100 bp
* 53221 53973: contig of 753 bp in length
* 53974 54073: gap of 100 bp
* 54074 54862: contig of 789 bp in length
* 54863 54962: gap of 100 bp
* 54963 55747: contig of 785 bp in length
* 55748 55847: gap of 100 bp
* 55848 56660: contig of 813 bp in length
* 56661 56760: gap of 100 bp
* 56761 57571: contig of 811 bp in length
* 57572 57671: gap of 100 bp
* 57672 58451: contig of 780 bp in length
* 58452 58551: gap of 100 bp
* 58552 59333: contig of 782 bp in length
* 59334 59433: gap of 100 bp
* 59434 60245: contig of 812 bp in length
* 60246 60345: gap of 100 bp
* 60346 61086: contig of 741 bp in length
* 61087 61186: gap of 100 bp
* 61187 61958: contig of 772 bp in length

Query Match      4.2%; Score 55.4; DB 2; Length 65523;
Best Local Similarity 36.7%; Pred. No. 0.036;
Matches 170; Conservative 0; Mismatches 293; Indels 0; Gaps 0;

QY 603 TAAATCCAAAGTAGGAGCGGTGCATGATGAGAATCGCTCAGTACTCGACATAATGAAC 662
DB 21280 TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 21221
QY 663 CTTACAGGTACACAGGCAGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 722
DB 21220 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 21161
QY 723 TGGCCCCACGGCGTGTGCGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 782
DB 21160 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21101
QY 783 CGCTGGATGCGCAAAATCAACAGAGGTGTGTCGCCCGCTGCGCTGCGCTGCGCTGCG 842
DB 21100 GGCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21041
QY 843 CATCCCTCTGCAGCGCTGCATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 902
DB 21040 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20981
QY 903 CCCTACTACGAAACCCCGTCCCGCGCGCGTGCCTGCGTGCCTGCGCTGCGCAAGAGGCG 962
DB 20980 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20921
QY 963 GCGCGCGGTGTGAGTCCCTGGACACCCGACACCTGTGCGGCGCTTTGTTTATTCATCC 1022
DB 20920 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20861
QY 1023 CGAATCTCTACTGCCCCCAGCGGCGACTGCGTGCCTGCGCGCGCGCGCGCGCGCG 1065
DB 20860 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 20818

RESULT 10
SGU36751
LOCUS
DEFINITION
Schizaphis graminum cytochrome oxidase subunit II (COII) gene,
mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION
U36751
VERSION
U36751.1 GI:1036830
KEYWORDS
greenbug.
ORGANISM
Mitochondrion Schizaphis graminum
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
```



E-mail: michael.bevan@bsrc.ac.uk  
 Information on performance of analysis and a more detailed  
 annotation of this entry and other sequences of chromosome 4 can be  
 viewed at: <http://websvr.mips.biochem.mpg.de/proj/thai/>.

FEATURES  
 source  
 1..89904  
 /organism="Arabidopsis thaliana"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 misc\_feature  
 1..11070  
 /note="overlap to BAC ; please refer to this entry for  
 analysis and annotation"  
 19417..19605  
 /gene="T25K17.10"  
 /number=1  
 19417..21056  
 /gene="T25K17.10"  
 join(19417..19605,19706..20001,20198..21056)  
 /gene="T25K17.10"  
 /EC\_number="4.4.1.14"  
 /note="strong similarity to ACC synthase , Malus  
 domestica, U73816  
 Contains Aminotransferases class-I pyridoxal-phosphate  
 attachmentsite, Aa\_Transfer\_Class\_1 [SLSKDLGLPGRVG]  
 contains EST gb:N65643, AA650853, Z33841"  
 /codon\_start=1  
 /product="1-aminocyclopropane-1-carboxylate synthase-like  
 protein"  
 /protein\_id="CAB38949.1"  
 /db\_xref="GI:4539416"  
 /translation="MGLPLMERSNNNNVLSRVAVSDTHGSDSPYPAGKAYDENP  
 YDESHNPSGVIQGLAENOVSFLLYLEKNPEGSMGSKGAPCFRENALFQDYHG  
 LKTFROMAFSFMQIRGKARPDPRIVLTAGTANELLTFLADPNALLVPIPIY  
 PGFDRDLRWTGVKIVPIHCDSSNHQIITPEALESATQARDANIRKRVLLTNPSNP  
 LGATVKQVLEGLDLCFCVRNHLVSDIYSGVSFHASSETVSAEIVENIDIVSKER  
 VHVYLSKDLGLPGRVGTIYSDNVVTRMSSTFLVSSQTHMLASLSDSEF  
 TEYIRINLRNRRRTDITVEGLKAGIECLKAGNGLFCWMNGLLEKTKDKGLQL  
 WDVILKELNRLSPGSSCHSCSEVGVFRVCFANSENTEIALKRIHEFMDRRRR"  
 19606..19705  
 /gene="T25K17.10"  
 /number=1  
 19706..20001  
 /gene="T25K17.10"  
 /number=2  
 20002..20197  
 /gene="T25K17.10"  
 /number=2  
 20198..21056  
 /gene="T25K17.10"  
 /number=3  
 26480..27228  
 /gene="T25K17.20"  
 join(26480..26661,27042..27228)  
 /gene="T25K17.20"  
 /note="strong similarity to predicted protein, Arabidopsis  
 thaliana, AC005169"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAB38950.1"  
 /db\_xref="GI:4539417"  
 /translation="NASKLLQLKACAEKFSVKHGTTYKQLLDKNKMYIOPATI  
 EKCNELSKQLLTRASIPGRSESEFWKVNHVKGLWKNRADLRKEDAGIALFLGLECF  
 AWYCAGEIVGRGFTFTGYYP"  
 26480..26661  
 /gene="T25K17.20"  
 /number=1  
 26662..27041  
 /gene="T25K17.20"  
 /number=1  
 27042..27228  
 /gene="T25K17.20"  
 /number=2

COMMENT  
 Information on performance of analysis and a more detailed  
 annotation of this entry and other sequences of chromosome 4 can be  
 viewed at: <http://websvr.mips.biochem.mpg.de/proj/thai/>.

FEATURES  
 source  
 1..89904  
 /organism="Arabidopsis thaliana"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="4"  
 misc\_feature  
 1..11070  
 /note="overlap to BAC ; please refer to this entry for  
 analysis and annotation"  
 19417..19605  
 /gene="T25K17.10"  
 /number=1  
 19417..21056  
 /gene="T25K17.10"  
 join(19417..19605,19706..20001,20198..21056)  
 /gene="T25K17.10"  
 /EC\_number="4.4.1.14"  
 /note="strong similarity to ACC synthase , Malus  
 domestica, U73816  
 Contains Aminotransferases class-I pyridoxal-phosphate  
 attachmentsite, Aa\_Transfer\_Class\_1 [SLSKDLGLPGRVG]  
 contains EST gb:N65643, AA650853, Z33841"  
 /codon\_start=1  
 /product="1-aminocyclopropane-1-carboxylate synthase-like  
 protein"  
 /protein\_id="CAB38949.1"  
 /db\_xref="GI:4539416"  
 /translation="MGLPLMERSNNNNVLSRVAVSDTHGSDSPYPAGKAYDENP  
 YDESHNPSGVIQGLAENOVSFLLYLEKNPEGSMGSKGAPCFRENALFQDYHG  
 LKTFROMAFSFMQIRGKARPDPRIVLTAGTANELLTFLADPNALLVPIPIY  
 PGFDRDLRWTGVKIVPIHCDSSNHQIITPEALESATQARDANIRKRVLLTNPSNP  
 LGATVKQVLEGLDLCFCVRNHLVSDIYSGVSFHASSETVSAEIVENIDIVSKER  
 VHVYLSKDLGLPGRVGTIYSDNVVTRMSSTFLVSSQTHMLASLSDSEF  
 TEYIRINLRNRRRTDITVEGLKAGIECLKAGNGLFCWMNGLLEKTKDKGLQL  
 WDVILKELNRLSPGSSCHSCSEVGVFRVCFANSENTEIALKRIHEFMDRRRR"  
 19606..19705  
 /gene="T25K17.10"  
 /number=1  
 19706..20001  
 /gene="T25K17.10"  
 /number=2  
 20002..20197  
 /gene="T25K17.10"  
 /number=2  
 20198..21056  
 /gene="T25K17.10"  
 /number=3  
 26480..27228  
 /gene="T25K17.20"  
 join(26480..26661,27042..27228)  
 /gene="T25K17.20"  
 /note="strong similarity to predicted protein, Arabidopsis  
 thaliana, AC005169"  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CAB38950.1"  
 /db\_xref="GI:4539417"  
 /translation="NASKLLQLKACAEKFSVKHGTTYKQLLDKNKMYIOPATI  
 EKCNELSKQLLTRASIPGRSESEFWKVNHVKGLWKNRADLRKEDAGIALFLGLECF  
 AWYCAGEIVGRGFTFTGYYP"  
 26480..26661  
 /gene="T25K17.20"  
 /number=1  
 26662..27041  
 /gene="T25K17.20"  
 /number=1  
 27042..27228  
 /gene="T25K17.20"  
 /number=2

exon  
 /gene="T25K17.30"  
 /number=1  
 28289..28369  
 /gene="T25K17.30"  
 /number=1  
 28289..29256  
 /gene="T25K17.30"  
 join(28289..28369,28417..28641,28748..28879,28966..29256)  
 /gene="T25K17.30"  
 /note="strong similarity to caffeoyl-CoA  
 O-methyltransferase, Pinus taeda, AF036095"  
 /codon\_start=1  
 /product="caffeoyl-CoA O-methyltransferase-like protein"  
 /protein\_id="CAB38951.1"  
 /db\_xref="GI:4539418"  
 /translation="MAKDEAKGLLKSEELYKVCPIYLIETSYVYPREPEVLRE  
 LRNITHNHPQAGMATPAGOLMGLMLNLVNAKRTIEGVETGYSLLLTALTLPEDGK  
 VTAIDMRNDSYEIGLPVKKAGVEHKIDFKESEALPALDELLNNKYNVEGGDFAFVDA  
 SKLNYWYHERLIRLLIKVGGIIVDNTLWGSVAEPDSDTPEWRIEVKATLELNKKL  
 SADRQVQISQAALGDGTTICRRLY"  
 28370..28416  
 /gene="T25K17.30"  
 /number=1  
 28417..28641  
 /gene="T25K17.30"  
 /number=2  
 28642..28747  
 /gene="T25K17.30"  
 /number=2  
 28748..28879  
 /gene="T25K17.30"  
 /number=3  
 28880..28965  
 /gene="T25K17.30"  
 /number=3  
 28966..29256  
 /gene="T25K17.30"  
 /number=4  
 29253..29604  
 /note="Pro-*tRNA*\_(TGG)"  
 /product="tRNA-Pro"  
 30136..30215  
 /gene="T25K17.40"  
 /number=1  
 30136..30663  
 /gene="T25K17.40"  
 join(30136..30215,30384..30663)  
 /gene="T25K17.40"  
 /note="strong similarity to ribosomal protein L31,  
 Nicotiana glutinosa, U23784  
 contains EST gb:H77007"  
 /codon\_start=1  
 /product="putative ribosomal protein"  
 /protein\_id="CAB38952.1"  
 /db\_xref="GI:4539419"  
 /translation="MSEKKRKEEVTVREYINLHRLHSCTEFKKAPKAKEIRKFA  
 EKEMGTQDVRVDVKNQIWSKGIKIRPRIRVARKNDDADEAKEEFSLSVTVAEI  
 PARELSGLGKTIIEED"  
 30216..30383  
 /gene="T25K17.40"  
 /number=1  
 30384..30663  
 /gene="T25K17.40"  
 /number=2  
 30910..31193  
 /note="EST AA586104 matches to position 30910-31193; no  
 obvious *genemodel*"  
 32558..33006  
 /gene="T25K17.50"  
 complement(32558..32659)  
 /gene="T25K17.50"  
 /number=1  
 complement(join(32558..32659,32745..32806,32898..33006))  
 /gene="T25K17.50"  
 complement(join(<32558..32659,32745..32806,32898..33006))

Unpublished  
3 (bases 1 to 82307)  
Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W.,  
Lemcke,K. and Mayer,K.F.X.  
Unpublished  
4 (bases 1 to 196286)  
EU Arabidopsis sequencing, project.  
Direct Submission  
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopfersplitz 18a, D-82152 Martinsried, FRG, E-mail:  
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
this fragment has an overlap with ATCHRIV63 at the 5' end and an  
overlap with ATCHRIV65 at the 3' end.

Location/Qualifiers  
1..196286  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
6687..8444  
/gene="AT4g25770"  
complement(6687..6833)  
/gene="AT4g25770"  
/number=1  
complement(join(6687..6833,6920..7036,7083..7163,  
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))  
/gene="AT4g25770"  
complement(join(6687..6833,6920..7036,7083..7163,  
7278..7370,7494..7733,7782..7882,7945..8042,8134..8444))  
/gene="AT4g25770"  
/notes="similarity to hypothetical protein T211.6 -  
Arabidopsis thaliana, PIR2:T00623"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB79432.1"  
/db\_xref="GI:7269428"  
/translacion="MTASFRRFQISPDLLIPSPRFLPPSQNQFRFLYLOFPAQPKN  
LQQEMWNRLIGICFPAANRKRLEITDNGGDFDADVWESAERPDHLVAVNGIV  
GCAADMKYAAEQVKKFPPKLVHRACFLTYRYCGSEASATLTFDGVKMGERLANE  
VVFQIVVSVGLVWHRSGGLKISFVAHSLGILVARIKLYEQDEYDSDSPSK  
EK SARGEIAGLEPMNPITFATPHLGSRGHRQFPLCGLPFLERTASQTAHLAAGRTG  
KHLFLIDMLLISSYDLNRSALNAKRVAVANRLSMVDFYFHLOEFALTNEKAMVGW  
RTSSIRPNELPKPNLATTDPNPHIYVVERGVNDNGSCQSTSTVVTVEQDTELEGPAL  
FLT"  
complement(6834..6919)  
/gene="AT4g25770"  
/number=1  
complement(6920..7036)  
/gene="AT4g25770"  
/number=2  
complement(7037..7082)  
/gene="AT4g25770"  
/number=2  
complement(7083..7163)  
/gene="AT4g25770"  
/number=3  
complement(7164..7277)  
/gene="AT4g25770"  
/number=3  
complement(7278..7370)  
/gene="AT4g25770"  
/number=4  
complement(7371..7493)  
/gene="AT4g25770"  
/number=4  
complement(7494..7733)  
/gene="AT4g25770"

FEATURES  
source  
gene  
exon  
gene  
CDS  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon





```

join(940. .1083,1286. .1597,1804. .2328,2426. .2653)
/gene="At2g15460"
/codon_start=1
/product="putative polygalacturonase"
/protein_id="A017390.2"
/db_xref="GI:16598543"
/translation="MACISFAFKALCFLEFFIVASRSTVPKPVNQRHGAKPDGKTF
DNAMATFISWRACKRISGSSKIYYPKGTFLYGGVEFGPCNKPTEIFIDTGLLAPAN
PSDIKQDTWNFRYIINLLSGSTGLDGQGSQHPNDHCNPCKIAMTGMFAFYVNF
NSNTKIDTILNSKMGHFNFSYVHNHNTIGVTITAPGDSPTNDGIGKMGSSNIQISDN
IGTGDDGIALSLGCTNLNLSNVHNGCPGHGSIYSGLSGSKDEKDKVLIVRDIFNGTS
DGRIKTWESSAKLILSVNFYVENITQMDVGPKNIDOKYCPHPPECHERKGESHVQI
QNKLLKNITGTSKNVAVNLQSKIFPCKNVELIDINIKQGVKDSGSTSVCENVVDGF
ARGKMFPRIV"
1409. .115321
/note="Sequence from clone F9013"
/complement(2357. .2378)
/rpt_family="(TAAAA)n"
3911. .4024
/rpc_family="(TAAAA)n"
join(<4336. .4479,4682. .4993,5202. .5726,5830. .>6057)
/gene="At2g15470"
44336. .>6057

```

```

CDS
join(4336..4479,4682..4993,5202..5726,5830..6050)
/gene="Atgsl5470"
/codon_start=1
/product="putative polygalacturonase"
/protein_id="AAD17391.1"
/db_xref="GI:4335713"
/translation="MACISFAFKALCFLEFIVVASTRTVRKVFNVQRHGA
DNAAFYSINSRACKRIVGSGSKIVPKGTFLGVGVGLKNP IEF I DGTI
PSDIKQDTWISNRYINHLISGSGTLDGQKQKSPHNDCHTNCPLAMTMA
NSNTKDTWISNRYINHLISGSGTLDGQKQKSPHNDCHTNCPLAMTMA
IGTDGDCIATLSTTNINSNVNGCPGHGIDVSGNLGNKDEKAKOILRDV
DGIRIKWESASAILVSNFYENQIMDVKGLNKINDDQKYPHPPCBHERKSI
LKLNIYGTSTKKNVAMNLQCKIPPCKNVELIDINIKNGVKDGSSTSVCE
GKMPFOHCLN"
repeat_region complement(5755..5776)
/rpt_family="TAAAA)n"
7220..7281
repeat_region complement(7698..7802)
/rpt_family="(TA)n"
repeat_region complement(7698..7802)

```

```
repeat_region      complement(7735...7838)
                    /rpt_family="(TAAA)n"
repeat_region      complement(8026...8142)
                    /rpt_family="(TAA)n"
repeat_region      complement(8080...8183)
                    /rpt_family="(GAAAA)n"
repeat_region      complement(8739...8780)
                    /rpt_family="(CAAAA)n"
mpna               to(n<8823...9813.9995...>10458)
```

gene  
/gene="At2g15480"  
<8823. .>10458

```

CDS
    /note="F9013.3"
    join(8823..9813,9995..10458)
    /gene="At2g15480"
    /codon_start=1

```

```

/gene="At2g15480"
/codon_start=1
/product="putative glucosyltransferase"
/protein_id="AA17392.1"
/acc="AF135714"
/ab_peg="AT_135714"

```

/translation - mkrvsvskrlgllfppcvgelgpcgcnadnfinsvqksdsdso  
/p.nakifekpibeaonqndpleigkifnppcvbelgpcgcnadnfinsvqksdsdso  
fukfufshkpmkqolessiettkfsalvadmfwpaatesaeklgvrlpvhgtsfss  
ccsynmrlhkpykqolessiettkfsalvadmfwpaatesaeklgvrlpvhgtsfss  
tmsfgvlynsfvelsavadfsvfakrwhieglfsglssqsfivwvnrknogind  
eclmldkctpgsvlyfsefgstnvlndwlelfaegfsglssqsfivwvnrknogind  
emlpgekerttqgllipgawoqfvlldhkaicgffthcgwnsa tgi taagi pmwt  
pmgawegnekllltvrlgrivngvntekmlvklrgk  
wakklgemaakavbevggssyndnvmfmeelngr"



*****	
* NOTE: Estimated insert size may differ from sequence length	
* (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ).	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 50 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
*****	
1	16332: contig of 16332 bp in length
16333	16432: gap of unknown length
16433	26891: contig of 10459 bp in length
26892	26891: gap of unknown length
26992	39114: contig of 12123 bp in length
39115	39214: gap of unknown length
39215	47706: contig of 8492 bp in length
47707	47806: gap of unknown length
47807	54905: contig of 7099 bp in length
54906	55005: gap of unknown length
55006	61730: contig of 6725 bp in length
61731	61830: gap of unknown length
61831	68922: contig of 7092 bp in length
68923	69022: gap of unknown length
69023	73955: contig of 4933 bp in length
73956	74055: gap of unknown length
74056	79739: contig of 5684 bp in length
79740	79839: gap of unknown length
79840	87203: contig of 7364 bp in length
87204	87303: gap of unknown length
87304	95716: contig of 8413 bp in length
95717	95816: gap of unknown length
95817	101220: contig of 5404 bp in length
101221	101320: gap of unknown length
101321	106991: contig of 5671 bp in length
106992	107091: gap of unknown length
107092	111184: contig of 4093 bp in length
111185	111284: gap of unknown length
111285	117002: contig of 5718 bp in length
117003	117102: gap of unknown length
117103	122719: contig of 5617 bp in length
122720	122819: gap of unknown length
122820	127676: contig of 4857 bp in length
127677	127776: gap of unknown length
127777	132248: contig of 4472 bp in length
132249	132348: gap of unknown length
132349	137419: contig of 5071 bp in length
137420	137519: gap of unknown length
137520	140519: contig of 3000 bp in length
140520	140619: gap of unknown length
140620	145613: contig of 4994 bp in length
145614	145713: gap of unknown length
145714	150743: contig of 5030 bp in length
150744	150843: gap of unknown length
150844	153229: contig of 2386 bp in length
153230	153329: gap of unknown length
153330	155705: contig of 2376 bp in length
155706	155805: gap of unknown length
155806	159641: contig of 3836 bp in length
159642	159741: gap of unknown length
159742	161937: contig of 2196 bp in length
161938	162037: gap of unknown length
162038	166136: contig of 4099 bp in length
166137	166236: gap of unknown length
166237	169654: contig of 3418 bp in length
169655	169754: gap of unknown length
169755	173209: contig of 3455 bp in length
173210	173309: gap of unknown length
173310	174872: contig of 1563 bp in length
174873	174972: gap of unknown length
174973	178370: contig of 3398 bp in length
178371	178470: gap of unknown length

*****	
* 178471 181766: contig of 3296 bp in length	
* 181767 181866: gap of unknown length	
* 181867 184177: contig of 2311 bp in length	
* 184178 184277: gap of unknown length	
* 186254 186353: gap of unknown length	
* 186354 186353: gap of unknown length	
* 186354 188889: contig of 2536 bp in length	
* 188890 188889: gap of unknown length	
* 188990 190730: contig of 1741 bp in length	
* 190731 190830: gap of unknown length	
* 190831 192237: contig of 1407 bp in length	
* 192238 192337: gap of unknown length	
* 192338 194466: contig of 2129 bp in length	
* 194467 194566: gap of unknown length	
* 194567 196224: contig of 1658 bp in length	
* 196225 196324: gap of unknown length	
* 196325 197794: contig of 1470 bp in length	
* 197795 197894: gap of unknown length	
* 197895 199371: contig of 1477 bp in length	
* 199372 199471: gap of unknown length	
* 199472 202264: contig of 2793 bp in length	
* 202265 202364: gap of unknown length	
* 202365 204373: contig of 2009 bp in length	
* 204374 204473: gap of unknown length	
* 204474 205726: contig of 1253 bp in length	
* 205727 205826: gap of unknown length	
* 205827 207670: contig of 1844 bp in length	
* 207671 207770: gap of unknown length	
* 207771 208820: contig of 1050 bp in length	
* 208821 208920: gap of unknown length	
* 208921 210151: contig of 1231 bp in length	
* 210152 211652: contig of 1401 bp in length	
* 211653 211752: gap of unknown length	
* 211753 212869: contig of 1117 bp in length	
* 212870 212969: gap of unknown length	
* 212970 214033: contig of 1064 bp in length.	
*****	
FEATURES	Location/Qualifiers
source	1. .214033
	/organism="Rattus norvegicus"
	/db_xref="taxon:10116"
	/clone="CH230-236F15"
BASE COUNT	63123 a 41277 c 42183 g 62289 t 5161 others
ORIGIN	
Query Match 4.1%; Score 53.6; DB 2; Length 214033;	
Best Local Similarity 50.8%; Pred. No. 0.12;	
Matches 128; Conservative 0; Mismatches 124; Indels 0; Gaps 0;	
QY 814	CCCGTGCCTCTCCAGTCCACGGCACCATCCTCTGCAGCGCTCAGCAGCCATGCC 873
Db 140327	CCCCGGAACCCCGTCCGCGCCCGCCCGCCCGCTTCTGCTCTCTGACCCCGCCCGCC 140268
QY 874	GTGTCGGGAACGCAACACCCCAACCACTACGAAACCCCTCCGCGCGTCC 933
Db 140267	TCCCTCCCCCCCCCTCCCGCCCGCCCGCCCGCTCTCTGCTCTCTGCGCCCGCC 140208
QY 934	CGTGTGCGTCCGCGTCCGCAACGAGGGCGCGCGCTGCTGAGTCCCTGGACACCGGA 993
Db 140207	CCCGTCTCTCCGCGCCCGCCCGCGCGCTCTCTGCGCGCTCTCTGCGCCCGCC 140148
QY 994	CACCTGTGCGCGCTTTTATTATTCATCCCGAAATTCATCTGCCCGCCCGCGCTGC 1053
Db 140147	TCCCTCCCGCGCCCTCCGCGCTTCCGCGCCCGCCCTCCGCGCTCCCTCCCGCTCC 140088
QY 1054	GCTGCGCGCGCC 1065
Db 140087	CCCGCGCGCGCC 140076

Search completed: November 7, 2002, 12:45:06  
Job time : 3522.86 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:29:53 ; Search time 166.436 Seconds  
(without alignments)  
13503.300 Million cell updates/sec

Title: US-09-905-558c-16  
Perfect score: 1309  
Sequence: 1 cccatgcgtcttctgtac.....caaggaaagtatcccatg 1309

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235	18.0	12313	21	AAZ35184
2	232.8	17.8	2126	24	AAS96571
3	124.6	9.5	1888	22	AAF81468
4	50.2	3.8	6327	22	AAS45444
5	50	3.8	10133	24	ABL32458
6	49.4	3.8	73334	24	ABL34124
7	49.2	3.8	11691	24	ABL34241
8	48.4	3.7	1337	20	AAZ17263
9	48	3.7	399	22	AAH93294

c 10	47.8	3.7	326	22	AAS56505	Human cDNA for an
c 11	47.8	3.7	73334	24	ABL34125	Human immune syste
c 12	47.4	3.6	6627	24	ABL33574	Human immune syste
c 13	47.2	3.6	6033	21	AA70152	Plasmodium falci
c 14	46.8	3.6	281	22	AA118482	Human breast cance
c 15	46.8	3.6	6107	24	AAS61315	Human gene regulat
c 16	46.6	3.6	20420	22	AAK73165	Human immune/haema
c 17	46.4	3.5	463	22	AAS56574	Human cDNA for an
c 18	46	3.5	6988	24	ABL34440	Human immune syste
c 19	45.8	3.5	2140	21	AAZ49411	Pea DRR206 protein
c 20	45.8	3.5	8591	24	AAS61295	Human gene regulat
c 21	45.6	3.5	3730	22	ABA19745	Human nervous syst
c 22	45.6	3.5	6699	24	AAS63321	Chemically pretrea
c 23	45.6	3.5	7049	24	ABL32157	Human immune syste
c 24	45.2	3.5	13326	24	ABL33712	Human immune syste
c 25	45.2	3.5	17918	24	AAS61418	Human gene regulat
c 26	45	3.4	687	24	ABL34283	Human immune syste
c 27	45	3.4	883	22	AA126436	Human breast cance
c 28	45	3.4	883	22	AA126587	Human breast cance
c 29	45	3.4	883	22	AA126703	Human breast cance
c 30	45	3.4	883	22	AA126709	Human breast cance
c 31	45	3.4	883	22	AA126724	Human breast cance
c 32	45	3.4	883	22	AA126756	Human breast cance
c 33	45	3.4	883	22	AA126774	Human breast cance
c 34	45	3.4	883	22	AA126785	Human breast cance
c 35	45	3.4	6478	22	AAS45417	Chemically pretrea
c 36	45	3.4	6626	22	AAS46810	Tumour suppressor
c 37	45	3.4	10329	24	ABL34123	Human immune syste
c 38	45	3.4	16217	24	ABL32625	Human immune syste
c 39	44.8	3.4	259	22	AA125325	Human breast cance
c 40	44.8	3.4	11473	24	ABL33355	Human immune syste
c 41	44.8	3.4	17419	22	AAS45393	Chemically pretrea
c 42	44.8	3.4	17419	24	ABL33295	Human immune syste
c 43	44.6	3.4	2286	20	AAZ24385	Arabidopsis farnes
c 44	44.6	3.4	5575	24	AAS61207	Human gene regulat
c 45	44.6	3.4	5975	24	AAS61096	Human gene regulat

## ALIGNMENTS

RESULT 1  
AAZ35184  
ID AAZ35184 standard; DNA; 12313 BP.  
XX  
AC AAZ35184;  
XX  
DT 13-MAR-2000 (first entry)  
XX  
DE Corn delta-12 desaturase fad2-2 gene 5' untranslated region.  
XX  
KW Corn; maize; transgenic plant; lipid; food; feedstuff;  
KW vegetable oil; seed oil; oleic acid; fatty acid desaturase;  
KW delta-12 desaturase; fad2-2; ss.  
XX  
OS Zea mays.  
XX  
FH Key  
FH intron  
FT Location/Qualifiers  
FT 5651..12301  
FT /\*tag= a  
FT TATA\_signal 5439..5444  
FT /\*tag= b  
XX  
PN WO9964579-A2.  
XX  
PD 16-DEC-1999.  
XX  
PF 09-JUN-1999; 99WO-US12884.  
XX  
PR 11-JUN-1998; 98US-0088987.  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX







[illegible]



Db	59115	CAATTAAAAAACAATAATATACAAACACACATTCCATTAACTATCAAAAAAATCATCA	59055
Qy	432	TAGTCTACAATGGAA	446
		I I I I I I I I	
Db	59055	TCATCTATCATATAA	59041
RESULT 7			
ABL34241/C			
ID	ABL34241	standard; DNA; 11691 BP.	
XX	AC	ABL34241;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	Human immune system associated gene SEQ ID NO: 2214.	
XX	KW	Human; immune system disease; cytosine methylation; antiasthmatic;	
KW	antiartherosclerotic; antianaemic; cytostatic; neoplastic;		
KW	neuroprotective; anti-HIV; anticoagulant; ophthalmological;		
KW	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;		
KW	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;		
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;		
KW	gene; ds.		
XX	OS	Homo sapiens.	
XX	XX		
PN	WO200200928-A2.		
XX	PN		
PD	03-JAN-2002.		
XX	XX		
PF	02-JUL-2001; 2001WO-EP07537.		
XX	PR		
PR	30-JUN-2000; 2000DE-1032529.		
XX	PR		
XX	01-SEP-2000; 2000DE-1043826.		
XX	XX		
PA	(EPIT-) EPIGENOMICS AG.		
XX	XX		
PI	Olek A, Piepenbrock C, Berlin K;		
XX	XX		
DR	WPI; 2002-130909/17.		
XX	XX		
PT	Nucleic acid comprising fragment of chemically modified gene, useful		
PT	for diagnosis and treatment of diseases associated with abnormal		
PT	cytosine methylation		
XX	XX		
PS	Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.		
XX	XX		
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel		
CC	diseases. The present sequence is a gene of the invention.		
XX	XX		
SQ	Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;		
Query Match 3.8%; Score 49.2; DB 24; Length 11691;			
Best Local Similarity 51.1%; Pred. No. 0.029;			
Matches 140; Conservative 0; Mismatches 133; Indels 1; Gaps 1;			
Qy	160	TGAATAGAGTCACTAATGCGTTAAGCATCGGTACTCCGTAGGTCACATTCTGTTCTTCT	219
Db	6938	TAAATATAACCTATTCTTAAATATATAAATAATATATACATAATATATTACCCA	6879
Qy	220	TATTTCTGCATATTTTATTTGTTGTTTACTGATTATACGAGTAGTTATACATCATGCAC	279
Db	6878	TAAITCTAAATACATCATATACATTTTAAATATAACTC-ATTAATAATATACAAAAAAT	6820
Qy	280	ATACATATCATCATATATCAATATTTTCTTAAATTAATTAACATAAAATGACT	339

CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.

XX Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;  
SQ Query Match 3.7%; Score 48.4; DB 20; Length 1337;  
Best Local Similarity 34.5%; Pred. No. 0.018;  
Matches 97; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 785 CTGGATCGGCAAAATCAAGAGAGTGGTGGCCCGTGGCCACTCTCCAGTCGACGGCACCA 844  
DB 381 CCNNNNCCNN 440  
QY 845 TCCTCTGCGAGCGCTCACAGCCATGCGGTGTCGCGGAAGGACACACACCCCAACC 904  
DB 441 NCCNCCCCCNNNCC 500  
QY 905 CACTCAGAAACCGCTCCCGCGGTGGCCGTGCGGTCCGGCTCGGCAAGGCGGC 964  
DB 501 CNCNCCCGCCCGCCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 560  
QY 965 CCGCGTGTGAGTCCCTGGACACCGGACAGCCTGTGCGGCCCTTGTATTATCATCCG 1024  
DB 561 CCCNCCNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 620  
QY 1025 AAATCTCATCTGCCCCCAGCGGAGTGGCGTGGCGCGGCC 1065  
DB 621 CCCNCCNCCCGCCCGCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 661

RESULT 9  
AAH93294/c  
ID AAH93294 standard; DNA; 399 BP.  
XX AC AAH93294;  
XX DT 04-OCT-2001 (first entry)  
XX DE Plasmodium falciparum MAL3P8 polynucleotide SEQ ID NO 16.  
XX KW Human; antisense-therapy; gene-therapy; diagnostic; forensic;  
XX KW gene mapping; ds.  
XX OS Plasmodium falciparum.  
XX PN WO200152616-A2.  
XX PD 26-JUL-2001.  
XX PF 22-DEC-2000; 2000WO-US35190.  
XX PR 23-DEC-1999; 99US-0471275.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-0552317.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Tang YT, Liu C, Drmanac RT;  
XX WI WIPI; 2001-451890/48.  
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
XX PT antibodies and research use -  
XX PS Example 4; Page 106; 135pp; English.  
XX CC The invention relates to an isolated human polynucleotide (AAH75398)  
CC encoding a novel polypeptide (AAG64527) useful in antisense-therapy and  
CC gene-therapy, in diagnostics, forensics, gene mapping and identification  
CC of mutations responsible for genetic disorders and other traits.  
CC Polynucleotide sequences with potential homology were also identified

CC (AAH93283-AAH93356).  
XX Sequence 399 BP; 225 A; 24 C; 15 G; 135 T; 0 other;  
SQ Query Match 3.7%; Score 48; DB 22; Length 399;  
Best Local Similarity 47.1%; Pred. No. 0.013;  
Matches 147; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 208 TTCTGTTCTTCTTATTTGTCATATTTTATTGTTTACTGATTATACGAGTAGTTAT 267  
DB 320 TTTTCTTTTATATATTTGTCATTTTATGTTGCAATAAACGATTATATATATCTTT 261  
QY 268 ACATACATGACATACATATCATCATATATTTTCTTAAATTAATAATAAA 327  
DB 260 TTTTACATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 201  
QY 328 CTAATAATGACTAAATTTCTAACACCAACGACATTTGTAATGTTTTCTCCAACAACTTTAC 387  
DB 200 ATATATATTTATTTATTTTACGATAGTTTAAATTTTTTTTATCTCGCAATTA 141  
QY 388 CTATTTCTACATTTGTTCTTATTTTTCGAATTTCACTCTATAAACACATAGTTCTACAAATGAAA 447  
DB 140 AACATTTTAAATTTGTTCTTATTTATTTATTAATTAATAATAATAATAATAATAATAATAATA 81  
QY 448 ACAGTGTCTTGTACGACTATACGCGATGTTGGCTTACAAACATAAGACAATATAGTCGT 507  
DB 80 TTGAATGTTTGTATATTTTAAAGGTTTTTTTTTTTTTTTTTTTTTTTGATACACAATTCATTA 21  
QY 508 TTGAAGATTGAA 519  
DB 20 TTTAAATATAA 9

RESULT 10  
AAS56505/c  
ID AAS56505 standard; cDNA; 326 BP.  
XX AC AAS56505;  
XX DT 18-DEC-2001 (first entry)  
XX DE Human cDNA for an ovarian cancer protein #129.  
XX KW Human; ss; ovarian cancer protein; cancer; tumour; ovarian cancer;  
XX KW endometrial cancer; cytostatic.  
XX OS Homo sapiens.  
XX PN WO200170976-A2.  
XX PD 27-SEP-2001.  
XX PF 20-MAR-2001; 2001WO-US09062.  
XX PR 21-MAR-2000; 2000US-190710P.  
XX PR 22-JUN-2000; 2000US-213748P.  
XX PR 19-DEC-2000; 2000US-257276P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Xu J, Pyle RA, Stolk JA;  
XX WI WIPI; 2001-607531/69.  
XX PT Nucleic acids encoding 222 polypeptides associated with ovarian and  
XX PT endometrial cancers, useful for diagnosing, preventing and treating  
XX PT cancers -  
XX PS Claim 1; Page 157; 187pp; English.  
XX CC The invention relates to human polynucleotides encoding proteins  
CC associated with ovarian and endometrial cancers. The polynucleotides and  
CC the proteins they encode may be used in the prevention, diagnosis and



XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-130909/17.  
DR  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX Claim 1; SEQ ID NO 1547; 32pp + Sequence Listing; German.  
PS  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC which are used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 6627 BP; 2052 A; 117 C; 1386 G; 3072 T; 0 other;  
SQ  
Query Match 3.6%; Score 47.4; DB 24; Length 6627;  
Best Local Similarity 49.2%; Pred. No. 0.066;  
Matches 155; Conservative 0; Mismatches 156; Indels 4; Gaps 1;  
QY 192 TACTCCGTAGGTACATCTGTTCTTCTTATTTGTCATATTTTATGTTCTTACTGA 251  
DB 6510 TACTCTTTAAATACCTACTAATCTCAATTCCTTTCATCCTTAACCTAACCTAAATAA 6451  
QY 252 TTATAGAGTAGTATPACATACATGACATACATATCATCATATATACATATATATTTT 311  
DB 6450 TTCAATATTATTAATTTACAATAAATAAACTAAACCAAAAAATTAATAACTTTA 6391  
QY 312 CTAATTT---AAATTAACACTAAATGACATAATTTCTACACCAACGACATGTAAT 367  
DB 6390 CTAACCTTTACCCAACTAAATAAATAAACTAAATTTTAAACCCAAAAAATAAACCAT 6331  
QY 368 GTTTTCTCCAACTTTTACCTATTCTAGATTGTTCTATTTTTCGAATTTCACTCTATAAC 427  
DB 6330 ATAATCTAAATAACCTAACTCCATATCATCTAAACATAAATCAATTAATCTT 6271  
QY 428 AACATAGCTACATGGAACAGTCTTTGTACGACTATATACGCGATGTGGCTACA 487  
DB 6270 ATAACAACCTCCAAACGCAACAATACTATCTCCATTTTCAAAACCAACAATAAACA 6211  
QY 488 ACATAAGACAAATATA 502  
DB 6210 CCAATCGAAAACTA 6196  
RESULT 13  
AAAT70152/C  
ID AAAT70152 standard; DNA; 6033 BP.  
XX  
AC AAAT70152;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.  
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoa; infection; insecticide; ds.  
XX Plasmodium falciparum.  
XX  
XX WO200025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
XX 05-NOV-1999; 99WO-US26796.  
XX  
XX 05-NOV-1998; 98US-0107131.  
PR

XX (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
XX Hoffman S, Carucci D, Gardner M, Venter JC;  
PI WPI; 2000-365347/31.  
XX  
XX Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX  
XX Disclosure; Page 493-495; 577pp; English.  
XX  
XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAAT70078 to AAAT70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 other;  
Query Match 3.6%; Score 47.2; DB 21; Length 6033;  
Best Local Similarity 49.6%; Pred. No. 0.071;  
Matches 121; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 205 ACATCTCTCTCTCTTATTTGTCATATTTTATTTGTTGTTTACTGATTAACAGTAGT 264  
DB 2523 ACAATCTTTTCTAUGTACAACCATATTTATATATATTTATATATATATAGTTACT 2464  
QY 265 TATACATACATGCACATACATATCATATATATACATATATTTTCTAAATTAATTA 324  
DB 2463 GTTAAATGTTATATCAATATTTTCAACATTTTATTTAAAAATTTAAATAATTTAAAAA 2404  
QY 325 AAACATAAAATGACTAAATTTCTACACCAACGACATTTGTTTCTCCAACTT 384  
DB 2403 TTTTGAAAAATTCATTCATATATATATTTTCTATATATTTGAAGATATATACACATCCA 2344  
QY 385 TACCTATTCTACATGTTCTTATTTTCGAATTTTCACTCTATAAACAACATAGTCTACAATGG 444  
DB 2343 TTCTTCTAATCTCTTTTTCATTTTGTGCTTTTGTGCTTTTATATATATTTAAAAAATTTG 2284  
QY 445 AAAA 448  
DB 2283 GAAA 2280  
RESULT 14  
AAL16482  
ID AAL16482 standard; cDNA; 281 BP.  
XX  
XX AAL16482;  
XX  
DT 07-DEC-2001 (first entry)  
XX

DE Human breast cancer expressed polynucleotide 8939.  
KW Human; breast cancer; cell marker; cytostatic; ss.  
XX Homo sapiens.  
XX WO200151628-A2.  
XX 19-JUL-2001.  
XX 10-JAN-2001; 2001WO-US00798.  
XX 14-JAN-2000; 2000US-0176077.  
XX 14-MAR-2000; 2000US-0189167.  
XX 24-MAR-2000; 2000US-0192099.  
XX 29-MAR-2000; 2000US-0193480.  
XX 15-MAY-2000; 2000US-0205230.  
XX 09-JUN-2000; 2000US-0211315.  
XX 25-JUL-2000; 2000US-0220534.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Lillie J, Xu Y, Wang Y, Steinmann K;  
XX WPI; 2001-451856/48.  
XX New peptide useful as a marker for the diagnosis of breast cancer -  
XX Claim 1; Page 1617; 3695pp; English.  
XX The invention relates to human breast cancer expressed polynucleotides  
XX (A107544-A126789) and methods of assessing whether a patient is  
XX afflicted with breast cancer by examining the correlation between the  
XX expression of certain markers and the cancerous state of breast cells.  
XX The polynucleotides and encoded polypeptides are potential markers for  
XX detecting, diagnosing, monitoring, characterising treating and  
XX potentially preventing breast cancer. The polynucleotides and encoded  
XX polypeptides are also useful for isolating compounds with cytostatic  
XX activity.  
XX Sequence 281 BP; 93 A; 32 C; 39 G; 117 T; 0 other;  
XX  
Query Match 3.6%; Score 46.8; DB 22; Length 281;  
Best Local Similarity 54.7%; Pred. No. 0.023;  
Matches 93; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 212 GTTCTCTTATTTGTCATATTTTATTTGTTTACTGATTATACAGTAGTTATACAT 271  
DB 9 GTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 68  
QY 272 ACATGCACATACATATCATATATATATATATATATATATATATATATATATATAT 331  
DB 69 ACATGCTTAATAATTAACCTTAAATTTATTTTAAAAATAAATTTATGTCATGAAACAA 128  
QY 332 AAATGACTAAATTTCTAACACCAACGACATTTGTAATGTTTCTCCACAA 381  
DB 129 TATGGCTTAAATTTGAAATTTCTATACATATGTAATTTATTTCTCCATAAAA 178  
RESULT 15  
AAS61315/C  
ID AAS61315 standard; DNA; 6107 BP.  
XX AAS61315;  
XX AC  
XX AC  
XX AC  
XX 29-JAN-2002 (first entry)  
XX Human gene regulation-associated gene oligonucleotide #270.  
XX Human; Gene regulation-associated gene; severe combined immunodeficiency;  
XX cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
XX asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
XX renal disease; Preeclampsia; cardiac allograft vascular disease;

KW colorectal cancer; thyroid cancer; oesophageal cancer; ds: tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX Homo sapiens.  
XX WO200177375-A2.  
XX 18-OCT-2001.  
XX 06-APR-2001; 2001WO-EP03968.  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-017470/02.  
XX New nucleic acid sequences from chemically modified genes associated  
XX with gene regulation, useful for analysing cytosine methylations for  
XX diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
XX disease -  
XX Disclosure; SEQ ID No 276; 26pp; English.  
XX The invention relates to 224 nucleic acid sequences comprising at least  
XX 18 bases of a chemically pretreated gene associated with gene regulation  
XX selected from 43 known genes (or complementary sequences). The  
XX chemical pretreatment converts cytosine bases unmethylated at the  
XX 5-position to uracil or another base with hybridisation behaviour  
XX dissimilar to cytosine, to enable analysis of cytosine methylations.  
XX The DNA sequences, oligomers (or sets/arrays) and method are  
XX useful in the diagnosis of diseases (or predisposition to diseases)  
XX associated with gene regulation and in therapy of such diseases, by  
XX enabling analysis of the cytosine methylation patterns of such genes,  
XX kits are provided. They are especially useful in diagnosis  
XX and therapy of e.g. severe combined immunodeficiency disease, cardiac  
XX disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
XX asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
XX preeclampsia, graft versus-host disease. The present sequence is a  
XX sequence included in the sequence data for this specification and is  
XX associated with the human gene regulation-associated genes.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 6107 BP; 1822 A; 187 C; 1337 G; 2761 T; 0 other;  
XX  
Query Match 3.6%; Score 46.8; DB 24; Length 6107;  
Best Local Similarity 49.2%; Pred. No. 0.091;  
Matches 123; Conservative 0; Mismatches 127; Indels 0; Gaps 0;  
QY 220 TATTTGTCATATTTTATTTGTTTACTGATTATACGAGTAGTTATACATATGAC 279  
DB 2615 TTTAT 2556  
QY 280 ATACATATCATCAT 339  
DB 2555 ATAAATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2496  
QY 340 AAATTTCTAACACCAACGACATTTGTTTCTCCAAACATTTTACCTATTCTACATT 399  
DB 2495 ATAAATATATATTAATTAATTAATTTTCATATCTTATTTACAAAACATTAATATCTA 2436  
QY 400 GTTCTATTTCGAATTTCACTCTATATAACACATAGTCTACAATGGAAACAGTGCTTTGT 459  
DB 2435 AAAAACAATAAATATCAATCAACAAACTATTTATTTTAAACAATTTATATACATACACATAAA 2376

us-09-905-558c-16.rng

Tue Nov 12 13:19:22 2002

QY 460 ACGACTATAT 469  
| | | | |  
Db 2375 AATTCTAAAT 2366

Search completed: November 7, 2002, 10:50:17  
Job time : 319.436 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:44:03 : Search time 38.9851 Seconds  
(without alignments)  
8247.617 Million cell updates/sec

Title: US-09-905-558c-16  
Perfect score: 1309  
Sequence: 1 cccatcgctgttcttac.....caagggaagtgcctccatg 1309

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.2	3.9	7218	1	US-08-232-463-14
2	43.4	3.3	2781	3	US-08-749-522-4
3	43	3.3	796	4	US-09-007-119-14
4	43	3.3	809	4	US-09-007-119-10
5	41.8	3.2	1890	6	5312912-3
6	41.6	3.2	1441	4	US-08-821-994-63
7	41	3.1	4254	2	US-08-443-639-7
8	40.6	3.1	6152	4	US-08-973-462-1
9	40.4	3.1	1000	1	US-08-599-252-96
10	40.4	3.1	1000	5	PCT-US96-06352-96
11	40.4	3.1	1000	5	PCT-US96-06353-96
12	40.4	3.1	10342	4	US-08-972-927-5
13	40	3.1	4673	1	US-07-638-431-1
14	40	3.1	4673	5	PCT-US92-00018-1
15	39.6	3.0	5829	4	US-09-004-838-109
16	39.2	3.0	51952	3	US-08-947-823-1
17	39.2	3.0	19124	2	US-08-487-828-13
18	39	3.0	320	4	US-09-165-264-11
19	38.8	3.0	923	4	US-09-004-731-66
20	38.8	3.0	923	4	US-08-749-699-66
21	38.6	2.9	320	4	US-09-165-264-13
22	38.4	2.9	1368	3	US-08-874-563-5
23	38.4	2.9	1368	3	US-08-577-483-14
24	38.4	2.9	1736	3	US-09-182-816-22
25	38.4	2.9	1736	3	US-09-182-816-24
26	38.4	2.9	1736	3	US-09-471-528-22
27	38.4	2.9	1736	3	US-09-471-528-24

Sequence 22, Appl  
Sequence 24, Appl  
Sequence 7, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 8, Appl  
Sequence 393, App  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl

## ALIGNMENTS

## RESULT 1

US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pT2gpt-Fls

Query Match

3.9%; Score 51.2; DB 1; Length 7218;





```

,
, LOCATION: (710)
, OTHER INFORMATION: (a or c or g or t/u)
,
, FEATURE:
,
, NAME/KEY: misc_feature
, LOCATION: (723)
, OTHER INFORMATION: (a or c or g or t/u)
,
, FEATURE:
,
, NAME/KEY: misc_feature
, LOCATION: (771)..(772)
, OTHER INFORMATION: (a or c or g or t/u)
,
, FEATURE:
,
, NAME/KEY: misc_feature
, LOCATION: (787)
, OTHER INFORMATION: (a or c or g or t/u)
,
, FEATURE:
,
, NAME/KEY: misc_feature
, LOCATION: (792)
, OTHER INFORMATION: (a or c or g or t/u)
,
, FEATURE:
,
, NAME/KEY: misc_feature
, LOCATION: (796)
, OTHER INFORMATION: (a or c or g or t/u)
,
, US-09-007-119-14

```

Query Match	3.3%;	Score 43;	DB 4;	Length 796;
Best Local Similarity	54.8%;	pred. No. 0.012;		

	Matches	85;	Conservative	0;	Mismatches	70;	Indels	0;	Gaps
Qy	221	ATTGTGTCACATTTTTTATGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACA	280						
Db	148	ATTGTGCAATTTGGTACTCCCTTGGACACGTTATATAGAAAATAATCAATTTTTCACA	207						
Qy	281	TACATATCATCATATATATACATATATTTTCTAAATTTAAATAAAAACTAAATAGCTA	340						
Db	208	TTAATTGAAGAAGTTAGTCTCACATCATTTAATTTTACTAATCTCAATAAAAAAATAGGT	267						
Qy	341	AATTTCTAACACCAACGACATTTGTAATGTTTCTC	375						
Db	268	TATTTCTAAAATGTCCTTCATTTGGAACCTGTTATC	302						

```

RESULT 4
US-09-007-119-10/c
; Sequence 10, Application US/09007119C
; Patent No. 6300541
; GENERAL INFORMATION:
; APPLICANT: Gibson, Paul T.
; APPLICANT: Markem, Khalid
; TITLE OF INVENTION: Soybean Sudden Death Syndrome Resistant Soybeans,
; TITLE OF INVENTION: Soybean Cyst Nematode Resistant Soybeans and Methods of
; TITLE OF INVENTION: Breeding and Identifying Resistant Plants
; FILE REFERENCE: Sou Illinois 1268/2 Sequence Listing
; Patent No. 6300541
; CURRENT APPLICATION NUMBER: US/09/007,119C
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,335
; EARLIER FILING DATE: 1997-01-14
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 809
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(6)
; OTHER INFORMATION: (a or c or g or t/u)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (506)
; OTHER INFORMATION: (a or c or g or t/u)
; FEATURE:

```

```

: NAME/KEY: misc_feature
: LOCATION: (619)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (562)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (687)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (747)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (766)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (769)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (785)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (795)
: OTHER INFORMATION: (a or c or g or t/u)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (797)
: OTHER INFORMATION: (a or c or g or t/u)
: US-09-007-119-10

```

Query Match 3.3%; Score 43; DB 4; Length 809;  
Best Local Similarity 54.8%; pred. No. 0.012;  
Matches 85; Conservative 0; Mismatches 70; Indels

Qy	221	ATTGTGCATATATTTTAACTGTTACTGATTATACAGTAGTTATACATACATCACACA	280
Db	457	ATTGTCAAAATTTGGTGACTCCCTTTGGACTCGTATATAAGAAAAATAACTAATTCACA	398
Qy	281	TACATATCATCATATATATACACAATATTTTTCTPAAATTAATTTAAAACATAAAATGACTA	340
Db	397	TTAATTAAAGAAGTTAGTTCACATCATTTAATTTTACTAATCCTCAAATAAAAAATAAGCT	338
Qy	341	AATTTCTAACACCACACGAATTTGTAATGTTTTCTC	375
Db	337	TATTTCTAAAATTCCTTTCAATGGAACCTGTTATC	303

RESULT 5  
5312912-3/C  
; Patent No. 5312912  
; APPLICANT: HADWIGER, LEE A.; CHIANG, CHIN C.; HOROVITZ,  
; DANIEL A.  
; TITLE OF INVENTION: PROCEDURES AND REGULATORY DNA SEQUENCES  
; FOR GENETICALLY ENGINEERING DISEASE RESISTANCE AND OTHER  
; INDUCIBLE TRAITS IN PLANTS  
; NUMBER OF SEQUENCES: 9  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/393,301  
; FILING DATE: 13-JUN-1989  
; SEQ ID NO:3  
; LENGTH: 1890  
5312912-3

Query Match 3.2%; Score 41.8; DB 6; Length 1890;  
Best Local Similarity 53.3%; Pred. No. 0.041;  
Matches 88; Conservative 0; Mismatches 77; Indels 0



; LENGTH: 6152  
 ; TYPE: DNA  
 ; ORGANISM: P.  
 US-08-973-462-1

Query Match 3.1%; Score 40.6; DB 4; Length 6152;  
Best Local Similarity 54.3%; Pred. No. 0.16;  
Matches 82; Conservative 0; Mismatches 69; Indels 0

Qy	199	TAGGTCACATTCGTGCTCTCTATTTTGGCATAATTTTTATTGTTGTTTACTGATTATACG	258
Db	6009	TAGCTCAAAATTTTTATCTCTCTAAATATATATACATATATATATATATATATATATATATAT	5950
Qy	259	AGTAGTTATACATACATGCATACATATCATCATATATATCAAAATATTTTTCTCAAATT	318
Db	5949	ATAAATTTTAATAAACACAAATAATTAATATACATATATCTGATCGAATAATATACAGT	5890

RESULT 9  
US-08-599-252-96  
; Sequence 96, Application US/08599252  
; Patent No. 5705343  
; GENERAL INFORMATION:  
; APPLICANT: DRAYNA, DENNIS T.  
; APPLICANT: FEDER, JOHN N.  
; APPLICANT: GNIERKE, ANDREAS  
; APPLICANT: KIMMEL, BRUCE E.  
; APPLICANT: THOMAS, WINSTON J.  
; APPLICANT: WOLFF, ROGER K.  
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
\* TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
; CITY: Washington  
; STATE: DC

Query Match 3.1%; Score 40.4; DB 1; Length 1000;  
Best Local Similarity 46.4%; Pred. No. 0.072;  
Matches 122; Conservative 0; Mismatches 141; Indels 0

QY 208 TTCTGTTCTCTTATTGTGTCATAATTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267

Db	117	TTATTTATTTATTTAAAAATAATTTAAATATAAATATTATTATTACTTATATTAT	176
Qy	268	ACATACATGCGACATACATATCATCATCACATATATCAATATTTTCTCAAAATTAATTA	337
Db	177	TTATTTATTAATAATTTATAATATGTCACATTATAAAATATTATTAAATTAATTA	236
Qy	328	CTAAAAATGACTAAATTTCTTAACCAACGACATGTGAATGTTTCTCCAAACAATTTAC	387
Db	237	TGTCACCTATATTTTAAATATCATAAATACATGAANAATACAAATTTTCTGTAGAATC	296
Qy	388	CTATTCACATTTGTTCTATTTCCGATTTCACTCTATAAACAACATAGTCTACAATG	447
Db	297	GTTTTCCTCTACTGTGAAATAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCN	356
Qy	448	ACAGTGCCTTTGTACGACTATATA	470
Db	357	TGTTGATAAATCATANNAANANA	379

RESULT 10  
PCT-US96-06352-96  
Sequence 96, Application PC/TUS9606352  
GENERAL INFORMATION:  
APPLICANT: DRAYNA, DENNIS T.  
APPLICANT: FEDER, JOHN N.  
APPLICANT: GIRMEL, ANDREAS  
APPLICANT: KIMMEL, BRUCE E.  
APPLICANT: THOMAS, WINSTON J.  
APPLICANT: WOLFF, ROGER K.  
TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY  
TITLE OF INVENTION: HEMOCHROMATOSIS  
NUMBER OF SEQUENCES: 124  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Ave. N.W., Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/06352  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,252  
FILING DATE: 09-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 9053-0001.21  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US96-06352-96

QY 208 TTCTGTTCTCTTATTTGTGCATATTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267

[illegible]

```

RESULT 11
PCT-US96-06583-96
; Sequence 96, Application PC/TUS96060583
; GENERAL INFORMATION:
; APPLICANT: DRAYNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIRKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: KIMMEL, WINSTON J.
; APPLICANT: THOMAS, ROGER K.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESS: MORRISON & FOEBSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06583
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,252
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US96-06583-96

```

[illegible]

```
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
QY 331 AAAATGACTAAATTTCTAACACCAAGACATGTAATGTTTCTCTCCACAACTTTACCTA 390
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10342 AAAAAGGAAATTTGATGCAATTTATATTTTAAAGTGTGTTTAAACAATGTTCAA 10283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 391 TTCTACATGTTCTATTTTCGAATTTCTACTCTATAAACAACATAGTCTACAATGGAACA 450
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10282 TTCTTATAGATATATGTTTAAATAGACATTTTTCGCAAAATGTCATGACAAACATG 10223
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 451 GTGCTTTGTACGACATATATACCGATGTGCTGCTACAACATAAGACAATATAGTCGTTTG 510
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10222 GTCCGTGTGAGAGAAATATATGTCATGTTGTCATGTCGCAAGTTGAAACCATGTTGG 10163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 511 AAGATTGAACCTAT 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10162 CATGTTGCACAT 10149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khushmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/638,431
; FILING DATE: 19910110
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avrom D.
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda gtl11-2-7 kb genomic expression
; CLONE: Py10.1111
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: 718..3195
; OTHER INFORMATION:
; US-07-638-431-1

Query Match 3.1%; Score 40; DB 1; Length 4673;
Best Local Similarity 52.4%; Pred. No. 0.21; Indels 0; Gaps 0;
Matches 88; Conservative 0; Mismatches 80;

QY 204 TACATTCTGTTCTTCTTATTATTGTCATATATTTTATTGTTGTTACTGATTATACGAGTAG 263
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3487 TTCATATTTTTTTAAACGATTTTAACTATTTTAACTATTTTAACTATTTGTCGTTA 3546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 264 TTATACATATATGCATATATATATATATATATATATATATATATATATATATATATAT 323
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3547 TAATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 324 AAAACTAAATGACTAAATTTCTAACACCAAGACATTTCTAATGTTT 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3607 AAAAAAAAAAAAAAAAAAACTTAAAGTTAATAACATTTTAGGTTT 3654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 14
PCT-US92-00018-1
; Sequence 1, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khushmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4673 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
; STRAIN: 17X(NL)
; DEVELOPMENTAL STAGE: erythrocytic stage
; TISSUE TYPE: Blood
; CELL TYPE: erythrocytic stage
; IMMEDIATE SOURCE:
; LIBRARY: Py-lambda gtl11-2-7 kb genomic expression
; FEATURE:
```

CLONE: Py10.1111  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 718..3195  
OTHER INFORMATION:  
PCT-US92-00018-1

Query Match 3.1%; Score 40; DB 5; Length 4673;  
Best Local Similarity 52.4%; Pred. No. 0.21;  
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 204 TACATCTCTCTCTTATTTGTCGATATTTTATTTGTTGTTTACTGATTATACGAGTAGTTATA 263  
Db 3487 TTATATATTTTAAACGATTTTTTTTAACTATTTTTTTTAACTAATTTCTCTCGTTA 3546  
QY 264 TTATACATCATGCACATACATATCATCATATATATCAATATTTTCTTAATTTAAATTT 323  
Db 3547 TATATATATATTTATATATCTCAATATTTTAAAGGTTACAAATTTCTTAAATATAAA 3606  
QY 324 AAACTAAATGACTAAATTTCTTACACCAACGACATGTATGTTT 371  
Db 3607 AAAAAAAAAAAAAAAAAAACTTAAAGTTAATAACATTTTGTAGTTT 3654

RESULT 15

US-09-004-838-109/c  
; Sequence 109, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:  
; APPLICANT: Michelmore, Richard W.  
; APPLICANT: Shen, Kathy  
; APPLICANT: Meyers, Blake  
; TITLE OF INVENTION: Procedures and Materials for  
; NUMBER OF INVENTION: Conferring Pest Resistance in Plants  
; NUMBER OF SEQUENCES: 140  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/004,838  
; FILING DATE: 09-JAN-1998  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/781,734  
; FILING DATE: 10-JAN-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Einhorn, Gregory P.  
; REGISTRATION NUMBER: 38,440  
; REFERENCE/DOCKET NUMBER: 023070-0788100S  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 109:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5829 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..5829 /note= "RG2K"  
; OTHER INFORMATION: /note= "RG2K"  
US-09-004-838-109

Query Match 3.1%; Score 40; DB 4; Length 5829;  
Best Local Similarity 45.9%; Pred. No. 0.24;  
Matches 136; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 209 TCTGTTCTCTTATTTGTCGATATTTTATTTGTTGTTTACTGATTATACGAGTAGTTATA 268  
Db 4460 TATATTTTGCCTTAATCTTCCAAATCTTATGAGGATTAGCCAACTGTGCTTTTCTTA 4401  
QY 289 CATACATGCACATACATATCATCATATATATCAATATTTTTCTTAAATTTAAATTTAAAC 328  
Db 4400 TAAAAATAATAATGAGTGATATCTTAACTAAAACTAGTTAATAAAAAAATTTAGTTACGATC 4341  
QY 329 TAAAAATGACTAAATTTCTTAACACACGACGACATTTGTTATGTTTCTCCACAACTTTTACC 388  
Db 4340 TATAACTGACATTTGGCTATTTTAAACATCTTTTAAAGCTTTATTTTATAAAATGATATAAT 4281  
QY 389 TATTTCTACATTTGTTCTTATTTTCGAATTTTCACCTCTATAAACAACATAGTCTACAATGGAAAA 448  
Db 4280 AAAGAAATCATTTATTTCTAGACATTTTAAATAAATAAAGGTTGGTTTACATTTATCA 4221  
QY 449 CAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACACATAGACAAATATAGT 504  
Db 4220 ATGACCATTTTAAATTTATATATATGCAATTTATGCAATTTAGTTCAAAAAATGGT 4165

Search completed: November 7, 2002, 10:48:43  
Job time : 86.9851 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 08:30:58 ; Search time 1237.03 Seconds  
(without alignments)  
14282.232 Million cell updates/sec

Title: US-09-905-558c-16

Perfect score: 1309

Sequence: 1 cccatcgctgttctgtctac.....caagggaagtgateccatg 1309

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_gss.\*  
13: em\_gss\_hum.\*  
14: em\_gss\_inv.\*  
15: em\_gss\_pln.\*  
16: em\_gss\_vrt.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	149.2	11.4	374	12	BH255115
C 2	123.8	9.5	223	12	A2921021
C 3	107.6	8.2	456	9	AW928053
C 4	58.2	4.4	1101	12	CNS0106X
C 5	57.8	4.4	500	9	AU087781
C 6	57.4	4.4	935	12	CNS006XK
C 7	56.6	4.3	534	12	BH255739
C 8	55.6	4.2	1058	12	CNS014D0
C 9	55.4	4.2	745	12	A2046300
C 10	55.2	4.2	776	12	AG130889
C 11	54.2	4.1	818	10	BI952390
C 12	54	4.1	798	12	AQ251130
C 13	54	4.1	1101	12	CNS00FVE
C 14	53.4	4.1	1101	12	CNS008WC
C 15	53.2	4.1	945	12	CNS04D0K
C 16	53	4.0	910	12	CNS006ON
C 17	53	4.0	914	12	A2674804

C 18	53	4.0	928	12	CNS000KY
C 19	53	4.0	936	10	BG852371
C 20	53	4.0	1101	12	CNS0039G
C 21	52.8	4.0	652	12	BH255629
C 22	52.8	4.0	928	12	CNS000KY
C 23	52.6	4.0	848	10	BI950718
C 24	52.6	4.0	1092	12	CNS020K7
C 25	52.6	4.0	1101	12	CNS017KE
C 26	52.6	4.0	1101	12	CNS017ZQ
C 27	52.4	4.0	479	10	BM163977
C 28	52.4	4.0	1101	12	CNS00G7F
C 29	52.2	4.0	914	12	CNS002JY
C 30	52.2	4.0	1101	12	CNS00H18
C 31	52	4.0	861	10	BI953364
C 32	52	4.0	942	9	AL536132
C 33	52	4.0	1059	12	CNS0155J
C 34	52	4.0	1101	12	CNS017EJ
C 35	51.8	4.0	802	12	CNS011T0
C 36	51.8	4.0	1043	12	CNS0145P
C 37	51.8	4.0	1101	12	CNS0023I
C 38	51.6	3.9	776	12	AG060484
C 39	51.6	3.9	969	12	CNS03PO0
C 40	51.4	3.9	553	9	AI657313
C 41	51.4	3.9	563	10	BG004199
C 42	51.4	3.9	925	10	BG441241
C 43	51.4	3.9	1101	12	CNS016LI
C 44	51.2	3.9	342	10	BM168596
C 45	51.2	3.9	358	10	BM168541

#### ALIGNMENTS

#### RESULT 1

BH255115/c

LOCUS BH255115 374 bp DNA linear GSS 29-NOV-2001  
DEFINITION RXLIBAM0001A01r Zea mays L. XLI methyl filtration maize root genomic shotgun library Zea mays genomic clone RXLIBAM0001A01r, DNA sequence.

ACCESSION BH255115

VERSION BH255115.1 GI:17150008

KEYWORDS GSS.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.

#### REFERENCE

1 (bases 1 to 374)

#### AUTHORS

Kim,S.W., Yu,Y., Lee,M.C., Yang,T.J., Main,D., Henry,D., Oates,R. and Wing,R.A.

#### TITLE

Genomic shotgun library from maize

#### JOURNAL

Unpublished (2001)

#### COMMENT

Contact: Wing RA

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: MATTAACTCTACTAAAGGG

Class: shotgun

High quality sequence stop: 374.

location/Qualifiers

1..374

/organism="Zea mays"

/strain="B73"

/db\_xref="taxon:4577"

/clone="RXLIBAM0001A01r"

/clone\_lib="Zea mays L. XLI methyl filtration maize root genomic shotgun library"

/tissue\_type="Root"

/lab\_host="XLIBLue"

/note="Vector: PCUGIBlu-1; Site\_1: Sau3AI;

FEATURES

source





BASE COUNT	203 a	62 c	107 g	124 t
ORIGIN				
Query Match	8.2%; Score 107.6; DB 9; Length 496;			
Best Local Similarity	78.0%; Pred.No. 2.4e-12;			
Matches 142;	Conservative 0; Mismatches 39; Indels 1; Gaps 1;			
QY	208	TTCGTCTCTTATTTGTCATATTTTATTGTTGGTTACTGATTATACGAGTAGTTAT	267	
DB	429	TTTTTCCCTCGTATTTGTCAT-GTTTCATGCTGTTTACTGCTTATGCAGTAGTTAT	371	
QY	268	ACATACATGCACATACATATCATCACATATATCACAATATTTTCTAAATTAATAAAA	327	
DB	370	GCACACATGCACATACATGTCATCACATATATCGCACATGTTTCTCGGATTAATAAAA	311	
QY	328	CTAAAAATGACTFAAATTTCTTAACCAACGACATGTAATGTTTCTCCACACTTTAC	387	
DB	310	CTAAAAATGCCCTAACTTTCTACAANAATCTAAAGTTTGTCGTCCAAACACATCAATAC	251	
QY	388 CT 389			
DB	250 AT 249			
RESULT 4				
CNS0106X/c				
LOCUS	CNS0106X 1101 bp DNA linear GSS 26-JUL-1999			
DEFINITION	Drosophila melanogaster genome survey sequence I7 end of BAC			
	BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly); genomic survey sequence.			
ACCESSION	AL098595			
VERSION	AL098595.1 GI:5610206			
KEYWORDS	GSS.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 1101)			
AUTHORS	Genoscope.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
COMMENT	- web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.			
FEATURES	Location/Qualifiers			
source	1..1101			
	/organism="Drosophila melanogaster"			
	/plasmid="pBelOBAC11"			
	/db_xref="taxon:7227"			
	/clone_lib="DrosBAC"			
	/clone="BACN03K20"			
	/note="end : T7"			
BASE COUNT	258 a	107 c	60 g	175 t
ORIGIN	501 others			
Query Match	4.4%; Score 58.2; DB 12; Length 1101;			
Best Local Similarity	19.2%; Pred.No. 0.05;			
Matches 87;	Conservative 182; Mismatches 184; Indels 0; Gaps 0;			
QY	209	TCGTCTCTTATTTGTCATATTTTATTGTTGTTTACTGATATACGAGTAGTTATA	268	
DB	1027	TTTTTHHTMTMTMTHTTMTMTHTTTTHHHHTMTMTMTHTTMTMTHTTMTHTTMM	968	
QY	269	CATACATGCACATATCATCATATATATTTTCTAAATTAATAAAA	328	

QY	280	ATACATATCATCATATATACAAATATTTTCTTAATTAATTAAATTAACAATAAAAAAGACT	339
Db	214	ATGTGTATTTTTTTTT	155
QY	340	AAATTTCTAACACCAGCAGATGTAATGTTTTCTCCACAACCTTTACCTATTCTACATT	399
Db	154	TATTAT-TAAATTTTAATTTAATCTAATGUGTGTTCTATATATATAATAATTANATAA	96
QY	400	GTTCTATTTCGAATTTCACTCTATAACACATAGTCTACAATGGAAAACAGTCGTTTTGT	459
Db	95	TTTTTTTATAAATNCGACGAAAGTGAAAAATGACTATTGTAAAAAATATTTCGGAT	36
QY	460	ACGACTATATA 470	
Db	35	ACATAAATACA 25	
RESULT 6			
CNS006XK/c			
LOCUS	CNS006XK	935 bp	DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #BACRI4N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL066051		
VERSION	AL066051		
KEYWORDS	GSS:		
SOURCE	AL066051.1 GI:4945019		
ORGANISM	fruit fly.		
REFERENCE	Drosophila melanogaster		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 935)		
JOURNAL	Genoscope.		
COMMENT	Direct Submission Submitted (02-JUN-1999). Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pietter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm. Location/Qualifiers 1 . 935 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACRI4N09" /note="end : T7"		
FEATURES			
source	1 . 935		
BASE COUNT	257 a	170 c	162 g 96 t 250 others
ORIGIN			
Query Match	4.4%; Score 57.4; DB 12; Length 935;		
Best Local Similarity	30.2%; Pred. No. 0.073;		
Matches 115; Conservative	86;	Mismatches 180;	Indels 0; Gaps 0;
QY	803	CGAGTGTGTGCCCGTGCACACTCTCCACGTCACCGACCATCCCTCTGCAGCGCTCA	862
Db	921	SGSSGCGSGCGSGSCCCCSCGCSGCCSCSSSSCCSCGSCGCGCGSC	862
QY	863	CCAGCGATGCCGTGTGCGGACGGCACACACCCCAACCCACTCAGGAACCCCGTC	922
Db	861	GGSSCCGCGCGSGSSGCGGCCGCGCGGCGSCCSSCGCGCGCGCGCGCG	802

Db	236	TTTCAGTTTTTCTTACTATCATGAGAAATATATATATATGTTGTTTACTATTAGAGTAG	295
Qy	264	TTATACATACATGCACATATCATCATATATATCAATAT-TTTTTCAATTAAT	322
Db	296	T-----ACACATGCAGTAATATGTCATCATATTTTATGCAATGTTTTTTGGATTAAT	350
Qy	323	TAAACTAAATGACTAAATTTTCAACCAACGACATTTGTTTCTTCCAAACAAC	382
Db	351	TAAAGCAATTAATTTGCTTAAATTTTCAACCAACTACATAAGATTTTTCATGTTTATAC	410
Qy	383	TTTACTATTTCTACATTTCTTATTTCCAAATTTCACTCTATAAACAACATAGTCTACAA	441
Db	411	TTGTCAAAGCTAGTTAAATAGGAGGTTATATAGTGTATAGGAAATATGGAATAA	469
RESULT 8			
LOCUS	CNS014D0	1058 bp DNA linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC		
	BACN1117 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL103998		
VERSION	AL103998.1	GI:5615609	
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1058)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
	- Web : www.genoscope.cns.fr)		
COMMENT	Determination of this BAC-end sequence was carried out as part of a		
	collaboration with the European Drosophila Genome Project (EDGP) -		
	http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC		
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre		
	d'Etude du Polymorphisme Humain) with funding provided by a MRC		
	project grant. The DNA was prepared from embryos by Alain Bucheton		
	and Genevieve Payan. It has been constructed in the vector		
	pBelOBAC11.		
FEATURES	Location/Qualifiers		
source	1..1058		
	/organism="Drosophila melanogaster"		
	/plasmid="pBelOBAC11"		
	/db_xref="taxon:7227"		
	/clone_lib="DrosBAC"		
	/clone="BACN1117"		
	/note="end : SP6"		
BASE COUNT	121 a	175 c	93 g
ORIGIN		396 t	273 others
Query Match	4.2%	Score 55.6;	DB 12; Length 1058;
Best Local Similarity	32.9%;	Pred. NO. 0.17;	
Matches	78;	Conservative 65;	Mismatches 94;
			Indels 0; Gaps 0;
Qy	186	ATCGGTTACTCCGTAGGTACATCTGTTCTTCTTATTTGTCATATTTTATGTTGT	245
Db	700	ATAGSCWATTCWSASTRATASMTSTSWTBTBWKRSTRATKAKBATCTVTWVTATA	759
Qy	246	TACTGATTATACAGTAGTTATACATACATACATATATATATATATATATAT	305
Db	760	HSMTSMWYTTACCMTSMCTSWTSTACCAWCYCTACCAWTATATCTMTATATACAC	819
Qy	306	ATTTTCTTAATTAATTAACCTAAAAATGACTAAATTTCTAACCAACGACATGTA	365
Db	820	ACATTSCTYMAHTPAYAHTTCACTWAAANAATATTSTAWATYMTTTSCTTSWVS	879
Qy	366	ATGTTTCTCCCAACCTTACCTATTTCTACATTTGTTCTTATTTTCGAATTTCTACT	422



XhoI; Morex (mla) plants were greenhouse grown in the R green lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of *Blumeria graminis* f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wang, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```
FEATURES
  source
    1..1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="RPCI-98"
    /clone="BACR32p18"
    /note="end : TET3"

BASE COUNT      267 a   223 c   31 g   184 t   396 others
ORIGIN
    Query Match      4.1%; Score 54; DB 12; Length 1101;
    Best Local Similarity 29.1%; Pred. No. 0.38;
    Matches 102; Conservative 89; Mismatches 159; Indels 0; Gaps 0;

QY 204 TACATTCGTCCTCTTATTTGTCATATTTTATTGTTTACTGATATACGAGTAG 263
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 551 TWAHAWHTMCAAAHTHTWATTAAWMMAYAAWATAAACCATATTTTAAATCCAAWAM 610
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 264 TTATACATACATGCATACATATCATCATATATCATCAATATTTTCTAAATTAAT 323
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 611 WTHMMWMCWMACTWATCCCTTTWAATHTTWTATMAAATWATTTWMTWTATY 670
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 324 AAAACTAAAATGACTAAATTTCTAACACCAACGACATGTATGTTTCTCCAACT 383
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 671 AHAATAATWMTTCWMTAMWTTATAAAWMAATWMAHTAWATAMATCTTANYTCT 730
      |:::| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 384 TTACCTATTCATATGTTCTATTTCGAATTTCACTCTATAAACACATAGTCTCAATG 443
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 731 ATCTCTTTAYAWATTTTYYATATTCACAATAAATWTCWAAMWMMAWMTWTAMCMMW 790
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 444 GAAACAGTGCTTTGACGACTATATACGCGATGTGTGCTACACATAGCAATATAG 503
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 791 CAATCTATTAATTAATCWTATYWWHWCTWTWTHHTHTHAYCHTHMACWHWCW 850

QY 504 TCGTTTGAAGATGAACCTATATATCGGTACGGTTAATCCGCTATGTAC 553
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 851 TCHTYCAAWACTWTAAADMMWTTCCCTCMTWTHAYCMTTWTATCWTG 900
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
CNS008WC      1101 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION      BACR18L14 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL052719
VERSION      AL052719.1 GI:4934268
KEYWORDS      GSS.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 1101)
REFERENCE      Genoscope.
AUTHORS      Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```
FEATURES
  source
    1..1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone_lib="RPCI-98"
    /clone="BACR18L14"
    /note="end : TET3"

BASE COUNT      145 a   171 c   122 g   340 t   323 others
ORIGIN
    Query Match      4.1%; Score 53.4; DB 12; Length 1101;
    Best Local Similarity 37.9%; Pred. No. 0.5;
    Matches 83; Conservative 44; Mismatches 92; Indels 0; Gaps 0;

QY 215 CTCCTATTGTGCATATTTTATTGTTTACTGATATACGAGTAGTTTACATACA 274
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 CGTTTTTANTTTNTTTTTTTTTTTTCAATYHYCCACCYAWAWATAAAATWMM 655
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 275 TCCACATACATATCATATATACATATATTTTCTAAATTAATTAATAACTAAAA 334
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 656 AAAATATATAACAAAYYYMAWYMAHAAAYCAAYATATAAAACAAWATAATAAT 715
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 335 TGACATAAATTTCTAACACCAACGACATGTTGTTTCTCCAACTTACCTATCT 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 716 AAAATWHTTTTATATAAWAHATWAAWATTTHTTTTWTAAATYTYWWWWT 775
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 395 ACATGTTCTATTTCGAATTTCACTCTATAAACACATA 433
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 776 TWATYTTTAATWMTTAWMTTYCTTAAATATAAWMWATA 814
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
CNS04DOK      945 bp      DNA      linear      GSS 21-MAY-2000
LOCUS      Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION      101H21 of library G from Tetraodon nigroviridis, genomic survey
              sequence.
ACCESSION      AL285149
VERSION      AL285149.1 GI:8023560
KEYWORDS      GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM      Tetraodon nigroviridis
              Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
              Tetraodontidae; Tetraodon.
              1 (bases 1 to 945)
REFERENCE      Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizes, C., Fisher, C.,
              Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
              Weissenbach, J.
              Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
              Unpublished
              2 (bases 1 to 945)
REFERENCE      Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
              Bernot, A., Fizes, C., Wincker, P., Brottier, P., Quetier, F.,
              Saurin, W. and Weissenbach, J.
              Human gene number estimate provided by genome wide analysis using
              Tetraodon nigroviridis DNA sequence
              Unpublished
              3 (bases 1 to 945)
REFERENCE      Genoscope.
AUTHORS      Direct Submission
TITLE      Direct Submission
```

Search completed: November 7, 2002, 12:25:35  
Job time : 1252.03 secs

